# JMIR Formative Research

Process evaluations, early results, and feasibility/pilot studies of digital and non-digital interventions
Volume 5 (2021), Issue 12 ISSN: 2561-326X Editor in Chief: Gunther Eysenbach, MD, MPH, FACMI


## Contents

### Original Papers


A Conversational Artificial Intelligence Agent for a Mental Health Care App: Evaluation Study of Its Participatory Design (e30053)
Morena Danieli, Tommaso Ciulli, Seyed Mousavi, Giuseppe Riccardi. . . . . 6

An Electronic Patient-Reported Outcome Mobile App for Data Collection in Type A Hemophilia: Design and Usability Study (e25071)
Francesco Petracca, Rosaria Tempre, Maria Cucciniello, Oriana Ciani, Elena Pompeo, Luigi Sannino, Valeria Lovato, Giancarlo Castaman, Alessandra Ghirardini, Rosanna Tarricone. . . . . 22

Use and Appreciation of a Web-Based, Computer-Tailored Diet and Physical Activity Intervention Based on the Self-determination Theory: Evaluation Study of Process and Predictors (e22390)
Juul Coumans, Anke Oenema, Catherine Bolman, Lilian Lechner. . . . . 34

Development of a Web-Based School Support System Within the AVATAR Project for Psychosocial Well-being in Adolescents: Pilot Feasibility Study (e24840)
Francesca Mastorci, Paolo Piaggi, Gabriele Trivellini, Cristina Doveri, Anselmo Casu, Luca Bastiani, Irene Marinaro, Cristina Vassalle, Alessandro Pingitore. . . . . 55

Development of the Shift Smartphone App to Support the Emotional Well-Being of Junior Physicians: Design of a Prototype and Results of Usability and Acceptability Testing (e26370)
Isabelle Counson, Alexandra Bartholomew, Joanna Crawford, Katherine Petrie, Geetanjali Basarkod, Victoria Moynihan, Josie Pires, Rachel Cohen, Nicholas Glozier, Samuel Harvey, Samineh Sanatkar. . . . . 65

The Use of Food Images and Crowdsourcing to Capture Real-time Eating Behaviors: Acceptability and Usability Study (e27512)
Katharine Harrington, Shannon Zenk, Linda Van Horn, Lauren Giurini, Nithya Mahakala, Kiarri Kershaw. . . . . 79

Users' Perceptions Toward mHealth Technologies for Health and Well-being Monitoring in Pregnancy Care: Qualitative Interview Study (e28628)
Jane Li, David Silvera-Tawil, Marlien Varnfield, M Hussain, Vanitha Math. . . . . 91

Initial Feasibility of the "Families Moving Forward Connect" Mobile Health Intervention for Caregivers of Children With Fetal Alcohol Spectrum Disorders: Mixed Method Evaluation Within a Systematic User-Centered Design Approach (e29687)
Christie Petrenko, Carson Kautz-Turnbull, Alicia Roth, Jennifer Parr, Cristiano Tapparello, Utku Demir, Heather Olson. . . . . 105

Conversational Agent for Healthy Lifestyle Behavior Change: Web-Based Feasibility Study (e27956)
Dhakshenya Dhinagaran, Thirunavukkarasu Sathish, AiJia Soong, Yin-Leng Theng, James Best, Lorainne Tudor Car. . . . . 129





XSL•FO
RenderX

Patient-Reported Outcome Measure for Real-time Symptom Assessment in Women With Endometriosis: Focus Group Study (e28782)
Esther van Barneveld, Arianne Lim, Nehalennia van Hanegem, Lisa Vork, Alexandra Herrewegh, Mikal van Poll, Jessica Manders, Frits van Osch, Wilbert Spaans, Gommert van Koeveringe, Desiree Vrijens, Joanna Kruimel, Marlies Bongers, Carsten Leue. . . . . 145

Assessing the Views of Professionals, Patients, and Care Partners Concerning the Use of Computer Tools in Memory Clinics: International Survey Study (e31053)
Aniek van Gils, Leonie Visser, Heleen Hendriksen, Jean Georges, Majon Muller, Femke Bouwman, Wiesje van der Flier, Hanneke Rhodius-Meester. . . . . 156

A Multimodal Messaging App (MAAN) for Adults With Autism Spectrum Disorder: Mixed Methods Evaluation Study (e33123)
Mohamad Hijab, Dena Al-Thani, Bilikis Banire. . . . . 170

Assessing Health Data Security Risks in Global Health Partnerships: Development of a Conceptual Framework (e25833)
Juan Espinoza, Abu Sikder, James Dickhoner, Thomas Lee. . . . . 184

Test-Retest Reliability of Home-Based Fitness Assessments Using a Mobile App (R Plus Health) in Healthy Adults: Prospective Quantitative Study (e28040)
I-I Lin, You-Lin Chen, Li-Ling Chuang. . . . . 199

Adoption and Appropriateness of mHealth for Weight Management in the Real World: A Qualitative Investigation of Patient Perspectives (e29916)
Jessica Breland, Khizran Agha, Rakshitha Mohankumar. . . . . 216

Creating a Digital Toolkit to Reduce Fatigue and Promote Quality of Life in Multiple Sclerosis: Participatory Design and Usability Study (e19230)
Sarah Thomas, Andy Pulman, Huseyin Dogan, Nan Jiang, David Passmore, Keith Pretty, Beth Fairbanks, Angela Davies Smith, Peter Thomas.
2 2 5

Mobile Intervention to Improve Sleep and Functional Health of Veterans With Insomnia: Randomized Controlled Trial (e29573)
Erin Reilly, Stephanie Robinson, Beth Petrakis, Melissa Gardner, Renda Wiener, Carmen Castaneda-Sceppa, Karen Quigley. . . . . 244

Tracking Lower Urinary Tract Symptoms and Tamsulosin Side Effects Among Older Men Using a Mobile App (PERSONAL): Feasibility and Usability Study (e30762)
Austin Lee, Stacey Kenfield, Elizabeth Wang, Anthony Enriquez, Akinyemi Oni-Orisan, Michael Steinman, Ida Sim, Benjamin Breyer, Scott Bauer. . . . . 260

An Alternative to the Light Touch Digital Health Remote Study: The Stress and Recovery in Frontline COVID-19 Health Care Workers Study (e32165)
Sarah Goodday, Emma Karlin, Alexandria Alfarano, Alexa Brooks, Carol Chapman, Rachelle Desille, Daniel Karlin, Hoora Emami, Nancy Woods, Adrien Boch, Luca Foschini, Mackenzie Wildman, Francesca Cormack, Nick Taptiklis, Abhishek Pratap, Marzyeh Ghassemi, Anna Goldenberg, Sujay Nagaraj, Elaine Walsh, Stress And Recovery Participants, Stephen Friend. . . . . 275

The Effect of an Automated Mobile Patient Engagement Application on Emergency Department Revisits: Prospective Observational Study (e17839)
Pothik Chatterjee, Adam Beck, Jenna Brager, Daniel Durand, Christopher D'Adamo. . . . . 290

Integration of Web Analytics Into Graduate Medical Education: Usability Study (e29748)
Jackson Massanelli, Kevin Sexton, Chris Lesher, Hanna Jensen, Mary Kimbrough, Anna Privratsky, John Taylor, Avi Bhavaraju. . . . . 298

Preventive Digital Mental Health for Children in Primary Schools: Acceptability and Feasibility Study (e30668)
Sian Davies, Jenni Jardine, Kerry Gutridge, Zara Bernard, Stephen Park, Tom Dawson, Kathryn Abel, Pauline Whelan. . . . . 306



XSL•FO
RenderX

The Life Goals Self-Management Mobile App for Bipolar Disorder: Consumer Feasibility, Usability, and Acceptability Study (e32450)
Kelly Ryan, Shawna Smith, Anastasia Yocum, Isabel Carley, Celeste Liebrecht, Bethany Navis, Erica Vest, Holli Bertram, Melvin McInnis, Amy Kilbourne. . . . . 317

User Reviews of Depression App Features: Sentiment Analysis (e17062)
Julien Meyer, Senanu Okuboyejo. . . . . 327

Privacy and Confidentiality Concerns Related to the Use of mHealth Apps for HIV Prevention Efforts Among Malaysian Men Who Have Sex With Men: Cross-sectional Survey Study (e28311)
Roman Shrestha, Celia Fisher, Jeffrey Wickersham, Antoine Khati, Rayne Kim, Iskandar Azwa, Colleen Mistler, Lloyd Goldsamt. . . . . 339

The Impact of Patient Characteristics on Their Attitudes Toward an Online Patient Portal for Communicating Laboratory Test Results: Real-World Study (e25498)
Rosian Tossaint-Schoenmakers, Marise Kasteleyn, Annelijn Goedhart, Anke Versluis, Esther Talboom-Kamp. . . . . 351

An Early Warning Mobile Health Screening and Risk Scoring App for Preventing In-Hospital Transmission of COVID-19 by Health Care Workers: Development and Feasibility Study (e27521)
Ronald Mbiine, Cephas Nakanwagi, Herve Lekuya, Joan Aine, Kawesi Hakim, Lilian Nabunya, Henry Tomusange. . . . . 360

Examining the Impact of Digital Components Across Different Phases of Treatment in a Blended Care Cognitive Behavioral Therapy Intervention for Depression and Anxiety: Pragmatic Retrospective Study (e33452)
Monica Wu, Robert Wickham, Shih-Yin Chen, Connie Chen, Anita Lungu. . . . . 370

A Patient-Reported Outcome Tool to Triage Total Hip Arthroplasty Patients to Hospital or Video Consultation: Pilot Study With Expert Panels and a Cohort of 1228 Patients (e31232)
Yvette Pronk, Peter Pilot, Walter van der Weegen, Justus-Martijn Brinkman, Berend Schreurs. . . . . 383

The Successful Synchronized Orchestration of an Investigator-Initiated Multicenter Trial Using a Clinical Trial Management System and Team Approach: Design and Utility Study (e30368)
Dinesh Mudaranthakam, Alexandra Brown, Elizabeth Kerling, Susan Carlson, Christina Valentine, Byron Gajewski. . . . . 393

Understanding Adolescents' Perceptions and Aspirations Towards Their Relationship With Personal Technology: Survey Study (e27852)
Muhammad Jee, Alsa Khan, Nazneen Nazneen. . . . . 404

A Digital Health Fall Prevention Program for Older Adults: Feasibility Study (e30558)
Claire Jacobson, Lauren Foster, Hari Arul, Amanda Rees, Randall Stafford. . . . . 413

Development of a Credible Virtual Clinician Promoting Colorectal Cancer Screening via Telehealth Apps for and by Black Men: Qualitative Study (e28709)
Danyell Wilson-Howard, Melissa Vilaro, Jordan Neil, Eric Cooks, Lauren Griffin, Taylor Ashley, Fatemeh Tavassoli, Mohan Zalake, Benjamin Lok, Folakemi Odedina, Francois Modave, Peter Carek, Thomas George, Janice Krieger. . . . . 424

Acceptability of an mHealth App for Youth With Substance Use and Mental Health Needs: Iterative, Mixed Methods Design (e30268)
Zachary Adams, Miyah Grant, Samantha Hupp, Taylor Scott, Amanda Feagans, Meredith Phillips, Kristina Bixler, Phani Nallam, Dorothy La Putt. . . . . 438

A Pragmatic Intervention Using Financial Incentives for Pregnancy Weight Management: Feasibility Randomized Controlled Trial (e30578)
Rebecca Krukowski, Brandi Johnson, Hyeonju Kim, Saunak Sen, Riad Homsi. . . . . 450



XSL•FO
RenderX

A Smartphone-Based Self-management Intervention for Individuals With Bipolar Disorder (LiveWell): Protocol Development for an Expert System to Provide Adaptive User Feedback (e32932)
Evan Goulding, Cynthia Dopke, Tania Michaels, Clair Martin, Monika Khiani, Christopher Garborg, Chris Karr, Mark Begale. . . . . 464

Detection of Clenbuterol-Induced Changes in Heart Rate Using At-Home Recorded Smartwatch Data: Randomized Controlled Trial (e31890)
Willem Elzinga, Samantha Prins, Laura Borghans, Pim Gal, Gabriel Vargas, Geert Groeneveld, Robert Doll. . . . . 480

Predicting Risk of Stroke From Lab Tests Using Machine Learning Algorithms: Development and Evaluation of Prediction Models (e23440)
Eman Alanazi, Aalaa Abdou, Jake Luo. . . . . 489

Using Artificial Neural Network Condensation to Facilitate Adaptation of Machine Learning in Medical Settings by Reducing Computational Burden: Model Design and Evaluation Study (e20767)
Dianbo Liu, Ming Zheng, Nestor Sepulveda. . . . . 499

Transitioning From In-Person to Remote Clinical Research on Depression and Traumatic Brain Injury During the COVID-19 Pandemic: Study Modifications and Preliminary Feasibility From a Randomized Controlled Pilot Study (e28734)
Lauren Fisher, Sylvie Tuchman, Andrew Curreri, Maggie Markgraf, Maren Nyer, Paolo Cassano, Grant Iverson, Maurizio Fava, Ross Zafonte, Paola Pedrelli. . . . . 515

Depression, Anxiety, and Daily Activity Among Adolescents Before and During the COVID-19 Pandemic: Cross-sectional Survey Study (e30702)
Anna Jolliff, Qianqian Zhao, Jens Eickhoff, Megan Moreno. . . . . 526

Pandemic Information Dissemination and Its Associations With the Symptoms of Mental Distress During the COVID-19 Pandemic: Cross-sectional Study (e28239)
Ole Amundsen, Asle Hoffart, Sverre Johnson, Omid Ebrahimi. . . . . 536

Re-examining COVID-19 Self-Reported Symptom Tracking Programs in the United States: Updated Framework Synthesis (e31271)
Miranda Janvrin, Jessica Korona-Bailey, Tracey Koehlmoos. . . . . 549

Impact of the COVID-19 Pandemic on a Physician Group's WhatsApp Chat: Qualitative Content Analysis (e31791)
Sawsan Abdel-Razig, Pascale Anglade, Halah Ibrahim. . . . . 559

Nursing Perspectives on the Impacts of COVID-19: Social Media Content Analysis (e31358)
Ainat Koren, Mohammad Alam, Sravani Koneru, Alexa DeVito, Lisa Abdallah, Benyuan Liu. . . . . 567

Patients' Perspectives on Qualitative Olfactory Dysfunction: Thematic Analysis of Social Media Posts (e29086)
Jane Parker, Christine Kelly, Barry Smith, Aidan Kirkwood, Claire Hopkins, Simon Gane. . . . . 585

Predictors to Use Mobile Apps for Monitoring COVID-19 Symptoms and Contact Tracing: Survey Among Dutch Citizens (e28416)
Stephanie Jansen-Kosterink, Marian Hurmuz, Marjolein den Ouden, Lex van Velsen. . . . . 592

The Plebeian Algorithm: A Democratic Approach to Censorship and Moderation (e32427)
Benjamin Fedoruk, Harrison Nelson, Russell Frost, Kai Fucile Ladouceur. . . . . 602



XSL•FO
RenderX

## Short Paper

Campus Smoking Policies and Smoking-Related Twitter Posts Originating From California Public Universities: Retrospective Study (e33331)
Joshua Yang, Raphael Cuomo, Vidya Purushothaman, Matthew Nali, Neal Shah, Cortni Bardier, Nick Obradovich, Tim Mackey. . . . . 508



XSL•FO
RenderX

Original Paper

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

Morena Danieli[1*], PsyD; Tommaso Ciulli[2], PhD; Seyed Mahed Mousavi[1*], MSc; Giuseppe Riccardi[1*], PhD

[1]Speech and Interactive Signal Lab, Department of Engineering and Computer Science, Università degli Studi di Trento, Trento, Italy

[2]IDEGO srl, Rome, Italy

[*]these authors contributed equally

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

## Abstract

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

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

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

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

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

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

**KEYWORDS**

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



XSL•FO
RenderX

## Introduction

### Background

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

An increasing number of review papers have studied the use of chatbots in mental health. Chatbots are an evolution of internet-mediated psychological interventions. Although the latter were developed for supporting psychological care by prescriptive models, chatbots aim to engage users in short *conversations* about their mental distress. In the mental health domain, chatbots are often based on scripted or *Eliza-style* dialogs [6,16]. Bendig et al [17] have analyzed the results from 10 pilot studies published between 2009 and 2018. The goal of these pilots was to assess user acceptance and effectiveness of the therapeutic recommendations, but many of them mostly included nonclinical samples. The meta-analysis by Bendig et al [17] supports the view that state-of-the-art mental health chatbots are still experimental and that little evidence for transferring results to real psychotherapy contexts is available. In addition, Lim and Penn [18], who studied the potential of the application of digital technology in schizophrenia therapy, have stressed the need for reliable data, and the recent review by Gaffney et al [19] has highlighted the need for relying on unbiased data. However, Gaffney et al [19] have also stressed the importance of focusing current research in this field on the identification of the key mechanisms of action of the conversational agent interventions. This is very important, and in our view this aspect may be improved by meeting 2 requirements; that is, on the one hand by basing the interaction model of conversational agents on principled theoretical explanations of psychological change and on the other hand by involving mental health professionals in the design studies of blended interventions. This paper takes both recommendations into careful consideration.

### Objective

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

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

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

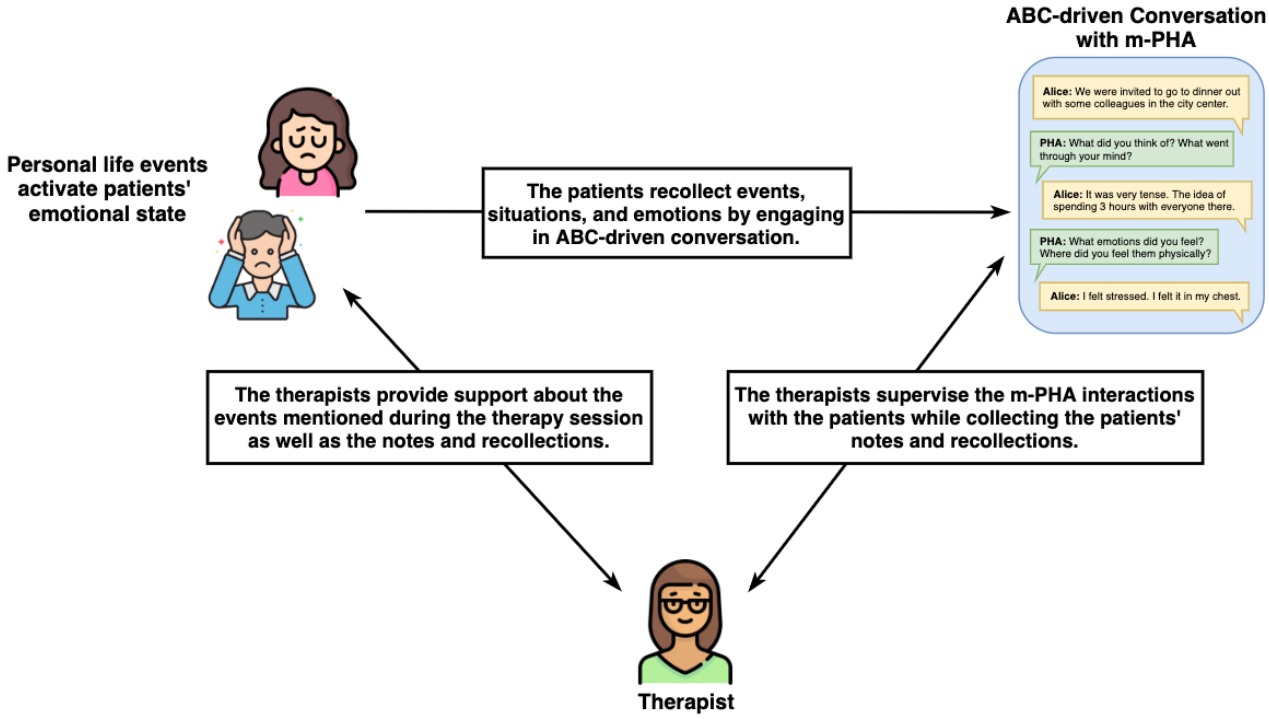
SMT programs are widely used for therapeutic purposes, with proven effectiveness. These programs combine specific techniques such as relaxation with CBT. This approach considers stress to be the imbalance between strong demands (external or internal) and few individual coping resources. The goal of SMT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

The approach is novel because it aims to (1) design the conversational features of the m-PHA to allow a natural and personal conversation and (2) allow the therapist to monitor patients' progress and difficulties during the time between a session and the one that follows. For this purpose, the m-PHA engages the patients in short conversations that are not scripted but are based on the recognition of their emotional state and on the understanding of the personal content written during the period of the intervention. For example, if the user reports issues in their relationship with colleagues—"Today was a bad day because my boss asked me to complete my assignment before the conveyed deadline"—the m-PHA asks contextually appropriate questions such as "You wrote that you had a bad day with your boss due to his request to finish a task in advance of the agreed time. What emotions did you have, what mental images and thoughts?" Figure 1 represents the information flow in the system architecture.



XSL•FO
RenderX

**Figure 1.** The information flow in the proposed approach. The patients interact with the mobile personal health care agent (m-PHA) to share personal recollections of their life events. The therapists supervise the interaction of the m-PHA with the patients and elaborate on the patients' personal narratives during the therapy session. ABC: antecedents, beliefs, and consequences; m-PHA: mobile personal health care agent.




A group of CBT therapists was involved in the process of designing this protocol as they provided information for identifying the variables that could be more suggestive of possible effectiveness of the approach. On the basis of these preliminary investigations, we set our research questions about the effectiveness (in terms of symptom reduction) of the joint use of psychotherapy and m-PHA, its possible persistence over time, and the acceptance of this integrated model by users and clinicians. The study is part of the European Union–funded Horizon 2020 research project COADAPT, whose aim is to develop methodologies to reduce work-related stress in aging workers.

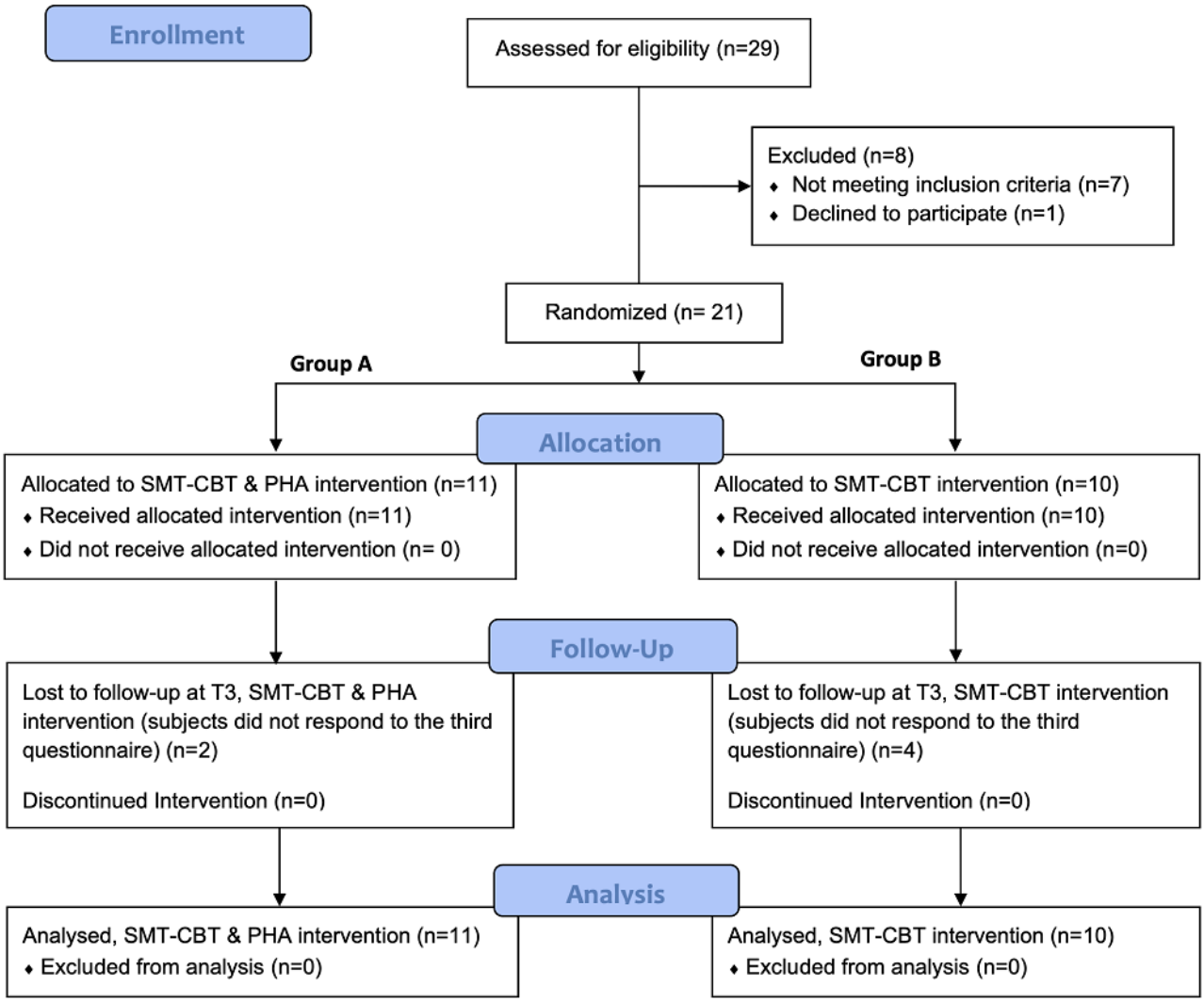
## Methods

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



XSL•FO
RenderX

**Figure 2.** The CONSORT (Consolidated Standards of Reporting Trials) diagram shows the flow of the intervention, the enrollment of participants, their allocation to treatment, follow-up, and analysis. CBT: cognitive behavioral therapy; m-PHA: mobile personal health care agent; SMT: stress management training; T3: assessment of psychopathological outcomes 3 months after treatment.




## Recruitment

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

## Description of the Questionnaires for Initial and Final Assessment

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



XSL•FO
RenderX

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

## Protocol

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

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

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

## Participants

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

**Table 1.** Sample characteristics (N=21).

| Characteristic | Values |
|---|---|
| Age (years), mean (SD) | 47.76 (8.07) |
| **Gender, n (%)** | |
| Male | 4 (19) |
| Female | 17 (81) |
| **Groups, n (%)** | |
| Group A | 11 (52) |
| Group B | 10 (48) |
| **Formal education, n (%)** | |
| High school | 7 (33) |
| Degree | 10 (48) |
| Master's degree or PhD[a] | 4 (19) |
| **Marital status, n (%)** | |
| Single | 3 (14) |
| Cohabiting | 4 (19) |
| Married | 13 (62) |
| Separated | 1 (5) |

[a]PhD: Doctor of Philosophy.



XSL•FO
RenderX

### Statistical Analysis

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

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

## Results

### Parametric Data Analysis

#### *Overview*

Parametric data analysis (independent *t* test) was performed on the collected data by comparing the differences between groups A and B with respect to the results obtained in the SCL-90-R, PSS (Table 2), and OSI (Table 3) questionnaires at T1, T2, and T3. For the OSI test, we only considered the scales regarding coping strategies such as home-work relationship, social support, logic, task oriented, involvement, and time.



XSL•FO
RenderX

**Table 2.** Parametric analysis of differences between group A (n=11) and group B (n=10) at baseline (T1), after 8 weeks of treatment (T2), and 3 months after treatment (T3): Perceived Stress Scale (PSS) and Symptom Checklist-90-Revised tests.

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

[a]GSI: Global Severity Index.

[b]PST: Positive Symptom Total.

[c]PSDI: Positive Symptom Distress Index.



XSL•FO
RenderX

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

| Scale | Group A, mean (SD) | Group B, mean (SD) | *t* test (*df*) | *P* value |
|---|---|---|---|---|
| **Social support** | | | | |
| T1 | 5.91 (1.51) | 5.90 (0.99) | 0.02 (19) | .99 |
| T2 | 6.36 (2.16) | 5.50 (1.35) | 1.08 (19) | .29 |
| T3 | 5.90 (2.18) | 7.33 (1.03) | –1.77 (13.6) | .10 |
| **Task oriented** | | | | |
| T1 | 5.45 (1.51) | 4.90 (1.37) | 0.88 (19) | .39 |
| T2 | 5.82 (1.33) | 6.10 (1.97) | –0.39 (19) | .70 |
| T3 | 5.30 (1.25) | 7.50 (1.05) | –3.60 (14) | .003 |
| **Logic** | | | | |
| T1 | 5.36 (1.96) | 4.80 (1.99) | 0.65 (19) | .52 |
| T2 | 5.55 (0.93) | 5.90 (1.52) | –0.65 (19) | .52 |
| T3 | 5.50 (1.58) | 5.67 (1.37) | –0.21 (14) | .83 |
| **Home-work relationship** | | | | |
| T1 | 5.45 (2.16) | 6.10 (1.79) | –0.74 (19) | .47 |
| T2 | 5.64 (1.96) | 6.20 (1.32) | –0.76 (19) | .45 |
| T3 | 6.10 (1.45) | 6.83 (1.17) | –1.05 (14) | .31 |
| **Time** | | | | |
| T1 | 5.00 (1.34) | 5.20 (1.99) | –0.27 (19) | .79 |
| T2 | 5.09 (1.87) | 5.80 (1.32) | –0.99 (19) | .33 |
| T3 | 5.10 (2.18) | 6.50 (2.07) | –1.26 (14) | .23 |
| **Involvement** | | | | |
| T1 | 5.73 (1.95) | 5.90 (1.10) | –0.25 (19) | .81 |
| T2 | 5.82 (1.08) | 5.60 (1.84) | 0.34 (19) | .74 |
| T3 | 6.00 (1.25) | 7.67 (1.37) | –2.5 (14) | .03 |

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

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

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

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

### *OSI Coping Strategies Results*

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

## Nonparametric Data Analysis

### *Overview*

Nonparametric data analysis (Mann–Whitney test) was performed on the collected data by comparing the differences between the groups with respect to the results obtained in SCL-90-R, PSS (Table 4), and OSI (Table 5) questionnaires at T1, T2, and T3. For the OSI test, we only considered the scales regarding coping strategies such as home-work relationship, social support, logic, task oriented, involvement, and time.



XSL•FO
RenderX

**Table 4.** Nonparametric analysis of differences between group A (n=11) and group B (n=10) at baseline (T1), after 8 weeks of treatment (T2), and 3 months after treatment (T3): Perceived Stress Scale (PSS) and Symptom Checklist-90-Revised tests.

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

[a]GSI: Global Severity Index.

[b]PST: Positive Symptom Total.



XSL•FO
RenderX

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

| Scale | Group A | | | Group B | | | *U* value | Z value | *P* value |
|---|---|---|---|---|---|---|---|---|---|
| | Mean (SD) | Median | Mean rank | Mean (SD) | Median | Mean rank | | | |
| **Social support** | | | | | | | | | |
| T1 | 5.91 (1.51) | 6.00 | 11.05 | 5.90 (0.99) | 6.00 | 10.95 | 54.5 | –0.04 | .99 |
| T2 | 6.36 (2.16) | 7.00 | 12.95 | 5.50 (1.35) | 6.00 | 8.85 | 33.50 | –1.56 | .13 |
| T3 | 5.90 (2.18) | 6.5 | 7.60 | 7.33 (1.03) | 7.00 | 10.00 | 21.00 | –1.00 | .32 |
| **Task oriented** | | | | | | | | | |
| T1 | 5.45 (1.51) | 5.00 | 12.05 | 4.90 (1.37) | 5.00 | 9.85 | 43.5 | –0.84 | .44 |
| T2 | 5.82 (1.33) | 6.00 | 10.14 | 6.10 (1.97) | 6.00 | 11.95 | 45.50 | –0.68 | .53 |
| T3 | 5.30 (1.25) | 6.00 | 5.95 | 7.50 (1.05) | 7.00 | 12.75 | 4.50 | –2.85 | .004 |
| **Logic** | | | | | | | | | |
| T1 | 5.36 (1.96) | 6.00 | 11.91 | 4.80 (1.99) | 5.00 | 10.00 | 45.00 | –0.73 | .49 |
| T2 | 5.55 (0.93) | 5.00 | 10.09 | 5.90 (1.52) | 6.00 | 12.00 | 45.00 | –0.73 | .5 |
| T3 | 5.50 (1.58) | 5.5 | 8.30 | 5.67 (1.37) | 6.00 | 8.83 | 28.00 | –0.22 | .85 |
| **Home-work relationship** | | | | | | | | | |
| T1 | 5.45 (2.16) | 5.00 | 10.05 | 6.10 (1.79) | 6.00 | 12.05 | 44.5 | –0.75 | .46 |
| T2 | 5.64 (1.96) | 6.00 | 10.27 | 6.20 (1.32) | 7.00 | 11.80 | 47.00 | –0.58 | .58 |
| T3 | 6.10 (1.45) | 6.00 | 7.60 | 6.83 (1.17) | 7.00 | 10.00 | 21.00 | –1.00 | .35 |
| **Time** | | | | | | | | | |
| T1 | 5.00 (1.34) | 6.00 | 10.86 | 5.20 (1.99) | 5.00 | 11.15 | 53.5 | –0.11 | .95 |
| T2 | 5.09 (1.87) | 6.00 | 9.50 | 5.80 (1.32) | 6.00 | 12.65 | 38.50 | –1.23 | .22 |
| T3 | 5.10 (2.18) | 5.00 | 7.40 | 6.50 (2.07) | 6.50 | 10.33 | 19.00 | –1.22 | .23 |
| **Involvement** | | | | | | | | | |
| T1 | 5.73 (1.95) | 6.00 | 10.23 | 5.90 (1.10) | 6.00 | 11.85 | 46.5 | –0.63 | .56 |
| T2 | 5.82 (1.08) | 6.00 | 11.09 | 5.60 (1.84) | 6.00 | 10.90 | 54.00 | –0.07 | .95 |
| T3 | 6.00 (1.25) | 6.00 | 6.55 | 7.67 (1.37) | 7.50 | 11.75 | 10.50 | –2.19 | .03 |

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

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

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

At T3, the depression level in group A (median 55) differed from that in group B (median 44.50; *U*=9.5; Z=–2.07; *P*=.04; *r*=–0.45). As for the PSS test, only at T3 did the levels reported by group A (median 18.5) differ significantly from those reported by group B (median 13; *U*=13.5; Z=–2.10; *P*=.03; *r*=–0.46).

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

### *OSI Coping Strategies Results*

For the subscales of the OSI test, the task-oriented level in group A (median 6) was significantly different from that in group B (median 7; *U*=4.5; Z=–2.85; *P*=.004; *r*=–0.62), and the involvement level in group A (median 6) was also significantly different from that in group B (median 7.5; *U*=10.5; Z=–2.19;



XSL•FO
RenderX

$P$=.02; $r$=–0.48) at T3. The analysis of the results at T1 and T2 reported in the OSI scale did not show other significant differences between group A and group B (Table 5). Participants assigned to group A reported lower OSI results than participants in group B in any subscale (higher levels are better), but only the task-oriented and involvement subscales significantly differed between the 2 groups at T3 (Table 5).

## Group A Within-Group Analysis

### *Parametric Data Analysis*

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

The level of PSS ($F_{2,18}$=3.25; $P$=.06) and some SCL-90-R subscales (GSI, $F_{2,16}$=2.80; $P$=.09; PST, $F_{2,16}$=1.58; $P$=.24; somatization, $F_{2,16}$=1.44; $P$=.27; interpersonal hypersensitivity, $F_{2,16}$=0.95; $P$=.41; depression, $F_{2,16}$=2.34; $P$=.13; anxiety, $F_{2,16}$=1.05; $P$=.37; hostility, $F_{2,16}$=0.43; $P$=.65; phobic anxiety, $F_{2,16}$=1.13; $P$=.35; paranoid ideation, $F_{2,16}$=1.26; $P$=.31; and psychoticism, $F_{2,16}$=1.47; $P$=.26) did not significantly change over the 3 measures at T1, T2, and T3.

For the PSDI and obsessiveness-compulsiveness subscales of the SCL-90-R test, the results show significant change over time (PSDI, $F_{2,16}$=6.47; $P$=.03, with moderate effect size $\eta^2$p=0.35 and obsessiveness-compulsiveness, $F_{2,16}$=6.58; $P$=.008, with large effect size $\eta^2$p=0.49).

The level of the examined OSI subscales of participants did not significantly change over the 3 measures at T1, T2, and T3 (social support, $F_{2,18}$=0.44; $P$=.65; task oriented, $F_{2,18}$=0.49; $P$=.62; logic, $F_{2,18}$=0.09; $P$=.92; home-work relationship, $F_{2,18}$=1.03; $P$=.37; time, $F_{2,18}$=0.04; $P$=.96; and involvement, $F_{2,18}$=0.22; $P$=.80).

### *Nonparametric Data Analysis*

A nonparametric data analysis was performed using the Friedman test (Pereira et al [39]) for comparing the different results reported in the participants in group A at T1, T2, and T3. The level of PSS (PSS, $\chi^2_2$=5.3; $P$=.07) and some SCL-90-R subscales (PST, $\chi^2_2$=4.2; $P$=.15; PSDI, $\chi^2_2$=4.2; $P$=.14; somatization, $\chi^2_2$=3.5; $P$=.20; interpersonal hypersensitivity, $\chi^2_2$=0.8; $P$=.71; depression, $\chi^2_2$=5.4; $P$=.08; anxiety, $\chi^2_2$=1.8; $P$=.45; hostility, $\chi^2_2$=0.8; $P$=.71; phobic anxiety, $\chi^2_2$=2.3; $P$=.33; paranoid ideation, $\chi^2_2$=1.7; $P$=.47; and psychoticism, $\chi^2_2$=3.0; $P$=.25) did not significantly change over the 3 measures at T1, T2, and T3.

The GSI and obsessiveness-compulsiveness subscales of participants significantly changed over the 3 measures at T1, T2, and T3 ($\chi^2_2$=6.4; $P$=.04; w=0.35 and $\chi^2_2$=6.4; $P$=.04; w=0.35, respectively). Wilcoxon tests were used to follow up this finding. A Bonferroni correction was applied; therefore, all effects have been reported at a 0.0167 level of significance. It seemed that the GSI did not significantly change from T1 to T2 ($t$=11.50; Z=–1.92; $P$=.06), from T1 to T3 (T=8; Z=–1.72; $P$=.09), or from T2 to T3 ($t$=13; Z=–0.169; $P$=.93). The obsessiveness-compulsiveness levels did not significantly change from T1 to T2 ($t$=11; Z=–1.96; $P$=.05) or from T2 to T3 ($t$=20; Z=–0.30; $P$=.79), but there was a significant change from T1 (median 63) to T3 (median 54; T=0.0; Z=–2.37; $P$=.02).

In summary, in group A, only the obsessiveness-compulsiveness levels showed a significant decrease at T1 in comparison with T3. The levels of the examined OSI subscales of participants did not significantly change over the 3 measures at T1, T2, and T3 (social support, $\chi^2_2$=0.6; $P$=.77; task oriented, $\chi^2_2$=1.3; $P$=.55; logic, $\chi^2_2$=0.7; $P$=.76; home-work relationship, $\chi^2_2$=0.9; $P$=.66; time, $\chi^2_2$=0.8; $P$=.71; and involvement, $\chi^2_2$=0.3; $P$=.90).

## Group B Within-Group Analysis

### *Parametric Data Analysis*

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

The level of PSS ($F_{2,12}$=3.56; $P$=.06) and some SCL-90-R subscales (PSDI, $F_{2,10}$=2.54; $P$=.17; obsessiveness-compulsiveness, $F_{1.1,5.48}$=2.86; $P$=.15; interpersonal hypersensitivity, $F_{2,10}$=3.85; $P$=.06; hostility, $F_{2,10}$=3.71; $P$=.06; phobic anxiety, $F_{2,10}$=0.147; $P$=.86; paranoid ideation, $F_{2,10}$=3.20; $P$=.08; and psychoticism, $F_{2,10}$=2.77; $P$=.11) did not significantly change over the 3 measures at T1, T2, and T3.

For the GSI scale, the Mauchly test indicated that the assumption of sphericity had been violated, $\chi^2_2$=7.1; $P$=.03; therefore, multivariate tests have been reported (ε=0.55). The results showed significant change over time, V=0.85, $F_{2,4}$=11.78; $P$=.02, with large effect size, $\eta^2$p=0.63.

For the PST, depression, somatization, and anxiety subscales of the SCL-90-R test, the results showed significant change over time (PST, $F_{2,10}$=8.87; $P$=.006; $\eta^2$p=0.64; depression, $F_{2,10}$=5.84; $P$=.02; $\eta^2$p=0.54; somatization, $F_{2,10}$=5.56; $P$=.02; $\eta^2$p=0.53; and anxiety, $F_{2,10}$=5.18; $P$=.03; $\eta^2$p=0.51).

The levels of the examined OSI subscales of participants did not significantly change over the 3 measures at T1, T2, and T3 (social support, $F_{2,10}$=3.89; $P$=.06; time, $F_{2,10}$=2.10; $P$=.17; and home-work relationship, $F_{2,10}$=3.57; $P$=.70).

For the task-oriented, logic, and involvement subscales of the OSI, the results showed significant change over time (task oriented, $F_{2,10}$=7.80; $P$=.009; $\eta^2$p=0.61; logic, $F_{2,10}$=9.54; $P$=.005; $\eta^2$p=0.66; and involvement, $F_{2,10}$=5.56; $P$=.02; $\eta^2$p=0.59).

### *Nonparametric Data Analysis*

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



XSL•FO
RenderX

The levels of PSS (PSS, $\chi^2_2$=4.5; *P*=.11) and some SCL-90-R subscales (PSDI, $\chi^2_2$=1.7; *P*=.52; somatization, $\chi^2_2$=5.0; *P*=.09; obsessiveness-compulsiveness, $\chi^2_2$=5.3; *P*=.07; interpersonal hypersensitivity, $\chi^2_2$=5.7; *P*=.06; anxiety, $\chi^2_2$=4.7; *P*=.11; hostility, $\chi^2_2$=3.9; *P*=.15; phobic anxiety, $\chi^2_2$=1.3; *P*=.59; paranoid ideation, $\chi^2_2$=4.3; *P*=.14; and psychoticism, $\chi^2_2$=4.3; *P*=.12) did not significantly change over the 3 measures at T1, T2, and T3. The GSI, PST, and depression subscales of participants significantly changed over the 3 measures at T1, T2, and T3 (GSI, $\chi^2_2$=9.5; *P*=.005; w=0.79; PST, $\chi^2_2$=9.0; *P*=.008; w=0.75; and depression, $\chi^2_2$=7.9; *P*=.01; w=0.66). Wilcoxon tests were used to follow up this finding. A Bonferroni correction was applied; therefore, all effects have been reported at a 0.0167 level of significance. It seemed that GSI did not significantly change from T1 to T2 (T=8; Z=–1.99; *P*=.04), from T2 to T3 (T=5.50; Z=–0.54; *P*=.69), and from T1 to T3 (T=0.0; Z=–2.23; *P*=.03).

PST did not significantly change from T1 to T2 (T=6; Z=–1.96; *P*=.05), from T2 to T3 (*t*=10.50; Z=0.0; *P*=.99), and from T1 to T3 (T=0.0; Z=–2.21; *P*=.03). Depression did not significantly change from T1 to T2 (*t*=10; Z=–1.79; *P*=.08), from T2 to T3 (T=9; Z=–0.31; *P*=.81), and from T1 to T3 (T=0.0; Z=–2.02; *P*=.06). The levels of the examined OSI subscales of participants did not significantly change over the 3 measures at T1, T2, and T3 (home-work relationship, $\chi^2_2$=0.5; *P*=.90 and time, $\chi^2_2$=3.4; *P*=.18). The social support, task-oriented, logic, and involvement subscales significantly changed over the 3 measures at T1, T2, and T3 ($\chi^2_2$=7.1; *P*=.03; $\chi^2_2$=8.5; *P*=.01; $\chi^2_2$=8.0; *P*=.01; and $\chi^2_2$=7.5; *P*=.01, respectively).

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

### Qualitative Evaluation of the Intervention

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

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

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

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

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

## Discussion

### Principal Findings

The analysis showed some significant differences between the 2 groups. The parametric analysis as well as the nonparametric analysis showed that in the examined subscales of the SCL-90-R, OSI, and PSS tests, group B seemed to show greater improvements than group A. The effect size in the parametric and nonparametric analyses was very large in scales that are significantly different.



XSL•FO
RenderX

In the SCL-90-R, for the subscales GSI, PST, anxiety, and depression, group B participants reported better changes on average than group A participants.

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

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

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

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

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

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

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

## Future Research

The goal of this study is to evaluate a protocol for an intervention for the treatment of work-related stress and anxiety based on the integration of a conversational AI-empowered mobile app into traditional psychotherapy. To validate the protocol, we needed to collect data from real users to feed the machine learning algorithms of the conversational m-PHA. More importantly, we needed to collect feedback from the psychotherapists who were involved in this participatory design effort. The limited number of participants that we could enroll did not allow the allocation of participants in more than 2 experimental groups. The research described in this paper was the initial and exploratory phase of a larger intervention protocol that is currently registered in ClinicalTrials.gov (NCT04809090). This larger protocol includes a control group, whose participants do not receive any type of treatment, as well as a fourth group, whose participants receive only the support of the m-PHA.

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

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

## Conclusions

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

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



XSL•FO
RenderX

## Acknowledgments

The research leading to these results has received funding from the European Union's Horizon 2020 program under grant agreement 826266: COADAPT.

## Conflicts of Interest

None declared.

## References

1. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep 10;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]
2. Thach K. User's perception on mental health applications: a qualitative analysis of user reviews. In: Proceedings of the 5th NAFOSTED Conference on Information and Computer Science (NICS). 2018 Presented at: 5th NAFOSTED Conference on Information and Computer Science (NICS); Nov. 23-24, 2018; Ho Chi Minh City, Vietnam p. 47-52. [doi: 10.1109/nics.2018.8606901]
3. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan 04;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
4. Solís-Galván J, Vázquez-Reyes S, Martínez-Fierro M, Velasco-Elizondo P, Garza-Veloz I, Caldera-Villalobos C. Towards development of a mobile application to evaluate mental health: systematic literature review. In: Proceedings of the International Conference on Software Process Improvement. 2020 Presented at: International Conference on Software Process Improvement; October 21-23, 2020; Mazatlán, Mexico p. 232-257. [doi: 10.1007/978-3-030-63329-5_16]
5. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO) - concept and usability test. IEEE Trans Emerg Topics Comput 2021 Jul 1;9(3):1170-1182. [doi: 10.1109/tetc.2020.2974478]
6. Bardram JE, Matic A. A decade of ubiquitous computing research in mental health. IEEE Pervasive Comput 2020 Jan;19(1):62-72. [doi: 10.1109/mprv.2019.2925338]
7. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. Usability, acceptability, and effectiveness of web-based conversational agents to facilitate problem solving in older adults: controlled study. J Med Internet Res 2020 Mar 12:e16794 [FREE Full text] [doi: 10.2196/16794] [Medline: 32384055]
8. Kilbourne AM, Beck K, Spaeth-Rublee B, Ramanuj P, O'Brien RW, Tomoyasu N, et al. Measuring and improving the quality of mental health care: a global perspective. World Psychiatry 2018 Feb 19;17(1):30-38 [FREE Full text] [doi: 10.1002/wps.20482] [Medline: 29352529]
9. Molodynski A, McLellan A, Craig T, Bhugra D. What does COVID mean for UK mental health care? Int J Soc Psychiatry 2020 Jun 10:20764020932592 [FREE Full text] [doi: 10.1177/0020764020932592] [Medline: 32517530]
10. Rossi R, Socci V, Talevi D, Mensi S, Niolu C, Pacitti F, et al. COVID-19 pandemic and lockdown measures impact on mental health among the general population in Italy. Front Psychiatry 2020 Aug 7;11:790 [FREE Full text] [doi: 10.3389/fpsyt.2020.00790] [Medline: 32848952]
11. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 1;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
12. Armontrout J, Torous J, Cohen M, McNiel D, Binder R. Exploring the boundaries of mental health computer applications. J Am Acad Psychiatry Law 2018;46(2):000-000 [FREE Full text]
13. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
14. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming state of practice. IEEE Access 2018;6:9390-9403. [doi: 10.1109/ACCESS.2018.2799522]
15. Rubeis G. E-mental health applications for depression: an evidence-based ethical analysis. Eur Arch Psychiatry Clin Neurosci 2021 Apr 01;271(3):549-555. [doi: 10.1007/s00406-019-01093-y] [Medline: 31894391]
16. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: a scoping review. Int J Med Inform 2019 Dec;132:103978. [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
17. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. Verhaltenstherapie 2019 Aug 20:1-13. [doi: 10.1159/000501812]
18. Lim M, Penn D. Using digital technology in the treatment of schizophrenia. Schizophr Bull 2018 Aug 20;44(5):937-938 [FREE Full text] [doi: 10.1093/schbul/sby081] [Medline: 29878251]
19. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
20. Matheny K, Gfroerer C, Harris K. Work stress, burnout, and coping at the turn of the century: an individual psychology perspective. J Indiv Psychol 2000;56(1):74 [FREE Full text]



XSL•FO
RenderX

21. Cilliers F. A systems psycho-dynamic perspective on burnout. SA J Ind Psychol 2003 Oct 24;29(4):26-33. [doi: 10.4102/sajip.v29i4.120]
22. Pines AM. The Female Entrepreneur: burnout treated using a psychodynamic existential approach. Clin Case Stud 2016 Jul 24;1(2):170-180. [doi: 10.1177/1534650102001002005]
23. Beck AT. Cognitive Therapy and the Emotional Disorders. Madison, CT: International Universities Press; 1976.
24. Ellis A. Reason and Emotion in Psychotherapy. Oxford, England: Lyle Stuart; 1962.
25. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 1;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
26. Blumenthal JA, Sherwood A, Smith PJ, Watkins L, Mabe S, Kraus WE, et al. Enhancing cardiac rehabilitation with stress management training. Circulation 2016 Apr 05;133(14):1341-1350. [doi: 10.1161/circulationaha.115.018926]
27. Thompson C, Poelmans S, Allen T, Andreassi J. On the importance of coping: a model and new directions for research on work and family. In: Perrewé PL, Ganster DC, editors. Exploring the Work and Non-work Interface (Research in Occupational Stress and Well Being, Vol. 6). Bingley: Emerald Group Publishing Limited; 2007:73-113.
28. Derogatis L. SCL-90: Administration Scoring and Procedures Manual-i and Other Instruments of the Psychopathology Rating Scale Series. Baltimore, MD: Johns Hopkins University School of Medicine; 1977.
29. Preti E, Prunas A, Sarno I, Madeddu F. SCL-90-R Symptom Checklist-90-R Adattamento italiano. Giunti: Organizzazioni Speciali; 2011.
30. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385. [doi: 10.2307/2136404]
31. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health: The Claremont Symposium on Applied Social Psychology. Thousand Oaks, CA: Sage Publications Inc; 1988:31-67.
32. Cooper C, Sloan S, Williams S. Occupational Stress Indicator: Management Set. Windsor: Nfer-Nelson; 1988.
33. Cooper CL, Williams J. A validation study of the OSI on a blue-collar sample. Stress Med 1991 Apr;7(2):109-112. [doi: 10.1002/smi.2460070210]
34. Conti S, Bonazzi S, Laiacona M, Masina M, Coralli MV. Montreal Cognitive Assessment (MoCA)-Italian version: regression based norms and equivalent scores. Neurol Sci 2015 Feb 20;36(2):209-214. [doi: 10.1007/s10072-014-1921-3] [Medline: 25139107]
35. Oltean H, Hyland P, Vallières F, David DO. An empirical assessment of REBT models of psychopathology and psychological health in the prediction of anxiety and depression symptoms. Behav Cogn Psychother 2017 Mar 28;45(6):600-615. [doi: 10.1017/s1352465817000133]
36. Sarracino D, Dimaggio G, Ibrahim R, Popolo R, Sassaroli S, Ruggiero GM. When REBT goes difficult: applying ABC-DEF to personality disorders. J Rat-Emo Cognitive-Behav Ther 2016 Dec 23;35(3):278-295. [doi: 10.1007/s10942-016-0258-7]
37. Whitfield G, Davidson A. Cognitive Behavioural Therapy Explained. Boca Raton, Florida, United States: CRC Press; 2018:8-11.
38. Sullivan G, Artino A. Analyzing and interpreting data from likert-type scales. J Grad Med Educ 2013 Dec;5(4):541-542 [FREE Full text] [doi: 10.4300/JGME-5-4-18] [Medline: 24454995]
39. Pereira DG, Afonso A, Medeiros FM. Overview of friedman's test and post-hoc analysis. Commun Stat Simul Comput 2014 Aug 21;44(10):2636-2653. [doi: 10.1080/03610918.2014.931971]
40. Berland A, Natvig GK, Gundersen D. Patient safety and job-related stress: a focus group study. Intensive Crit Care Nurs 2008 Apr;24(2):90-97. [doi: 10.1016/j.iccn.2007.11.001] [Medline: 18096388]

## Abbreviations

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



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 29.04.21; peer-reviewed by P Catellani, M Bennion, B Sindelar; comments to author 30.05.21; revised version received 20.07.21; accepted 19.09.21; published 01.12.21.*

*Please cite as:*
*Danieli M, Ciulli T, Mousavi SM, Riccardi G*
*A Conversational Artificial Intelligence Agent for a Mental Health Care App: Evaluation Study of Its Participatory Design*
*JMIR Form Res 2021;5(12):e30053*
*URL: https://formative.jmir.org/2021/12/e30053*
*doi:10.2196/30053*
*PMID:34855607*

©Morena Danieli, Tommaso Ciulli, Seyed Mahed Mousavi, Giuseppe Riccardi. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# An Electronic Patient-Reported Outcome Mobile App for Data Collection in Type A Hemophilia: Design and Usability Study

Francesco Petracca[1], MSc; Rosaria Tempre[2], MSc; Maria Cucciniello[1,3], PhD; Oriana Ciani[1,4], PhD; Elena Pompeo[2], MSc; Luigi Sannino[2], MSc; Valeria Lovato[2], PhD; Giancarlo Castaman[5], MD; Alessandra Ghirardini[2], MSc; Rosanna Tarricone[1,6], PhD

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

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

## Abstract

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

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

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

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

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



XSL•FO
RenderX

**Trial Registration:** ClinicalTrials.gov NCT04165135; https://clinicaltrials.gov/ct2/show/NCT04165135

***(JMIR Form Res 2021;5(12):e25071)*** doi:10.2196/25071

**KEYWORDS**

mobile apps; mHealth; hemophilia A; rare diseases; usability; user-centered design; design science; mobile phone

# Introduction

## Background

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

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

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

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

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

## Objectives

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

This effect can be partly attributed to the lack of stakeholder input in the development of mHealth technologies: the apps are often made available to the public without sufficient attention devoted to their design [24] and without a thorough understanding of the context of their proposed deployment and the needs of their end users, regardless of whether they are patients, caregivers, or clinicians [25].

Therefore, to maximize electronic outcome reporting and to ensure continuous data collection throughout the POWER study, we incorporated users' expectations, experiences, and needs in the design process of the ePRO app. This paper reports on the process adopted for the development of a mobile app aimed at collecting PROMs in patients with hemophilia A during the POWER study.

# Methods

## Theoretical Framework: User-Centered Design Approach

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

According to the UCD approach, during the concept generation phase, a thorough investigation of the needs is conducted to understand the environment of the end users and their requirements. On the basis of this investigation, a set of



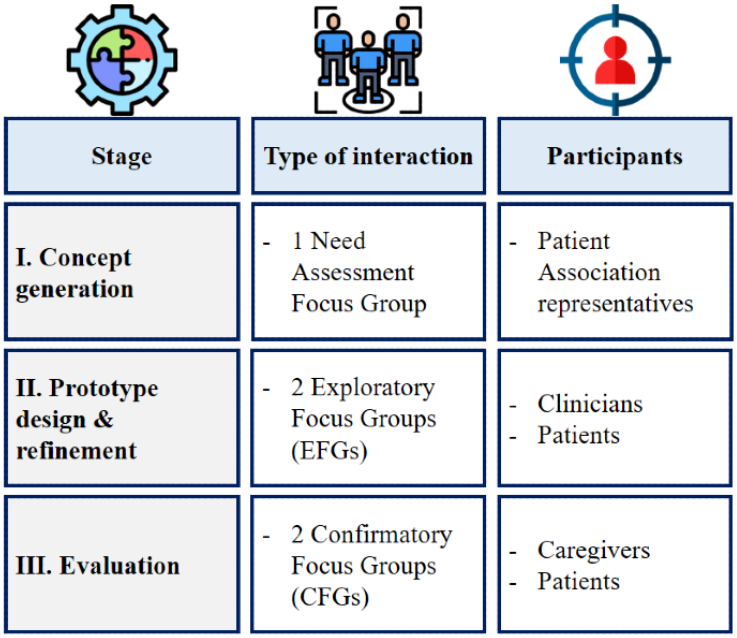
XSL•FO
RenderX

functional requirements were identified for the ePRO app through thematic analysis, leading to the design of a prototype. The next phases of the design cycle included continuous, iterative evaluation and refinement of the prototype and its usability testing.

## Study Design

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

**Figure 1.** Design process of the electronic patient-reported outcome app.




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

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

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

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

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

After the test, the participants were asked to quantitatively assess the usability of the app, a multidimensional property associated with attributes such as ease of use, user satisfaction, attractiveness of the layout, and error rates compared with the intended use [30]. According to the World Health Organization, usability testing constitutes the initial step of the evaluation process of any digital health technology [31]. To evaluate the usability of the ePRO app, participants completed the mHealth app usability questionnaire (MAUQ), an 18-item questionnaire recently developed by scholars on a 7-point Likert scale ranging from 1 (*strongly disagree) to 7 (strongly agree)* [32]. The MAUQ is the first scale developed to specifically assess the usability of mHealth apps, and its reliability and validity were



XSL•FO
RenderX

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

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

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

## Participants

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

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

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

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

## Data Analysis

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



XSL•FO
RenderX

## Results

### Participant Characteristics

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

**Table 1.** Participant characteristics (N=22).

| Characteristics | Need-assessment FG[a], n (%) | EFGs[b], n (%) | CFGs[c], n (%) | Total, n (%) |
|---|---|---|---|---|
| **Category** | | | | |
| Caregiver | 0 (0) | 0 (0) | 3 (33) | 3 (14) |
| Clinician | 0 (0) | 4 (44) | 0 (0) | 4 (18) |
| Patient | 0 (0) | 5 (56) | 6 (67) | 11 (50) |
| Patient representative | 4 (100) | 0 (0) | 0 (0) | 4 (18) |
| **Age (years)** | | | | |
| <30 | 0 (0) | 4 (44) | 4 (44) | 8 (36) |
| 30-39 | 0 (0) | 1 (11) | 1 (11) | 2 (9) |
| 40-49 | 1 (25) | 0 (0) | 1 (11) | 2 (9) |
| 50-59 | 1 (25) | 2 (22) | 1 (11) | 4 (18) |
| >60 | 2 (50) | 2 (22) | 2 (22) | 6 (27) |
| **Gender** | | | | |
| Male | 3 (75) | 6 (67) | 7 (78) | 16 (73) |
| Female | 1 (25) | 3 (33) | 2 (22) | 6 (27) |

[a]FG: focus group.

[b]EFG: exploratory focus group.

[c]CFG: confirmatory focus group.

### Concept Generation: Need-Assessment FG

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

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

> *Rather than improving the quality and quantity of the app experience, I believe it is fundamental to make it easy, simple, user-friendly, and self-explanatory, making it possible to access it also for somebody who would not want to know how it actually works.* [Patient Representative, 53 years]

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

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

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

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

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

The aesthetics of the user interface were identified as another factor to improve, with the FG participants advising to adopt different color sets to make the layout catchier, but not trivial, than the initial one and to enable individuals to tailor some of the app's graphical features to their personal preferences.



XSL•FO
RenderX

One final emerging theme pertained to the type of device used during the POWER study. The participants unanimously agreed that providing them with an additional device exclusively for study purposes would prove excessively burdensome and would not help maximize patient engagement, probably generating increased attrition rates instead. One patient representative stated as follows:

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

## Changes to the App After Need-Assessment

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

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

## Prototype Refinement: EFGs

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

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

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

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

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

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

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

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

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

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

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

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

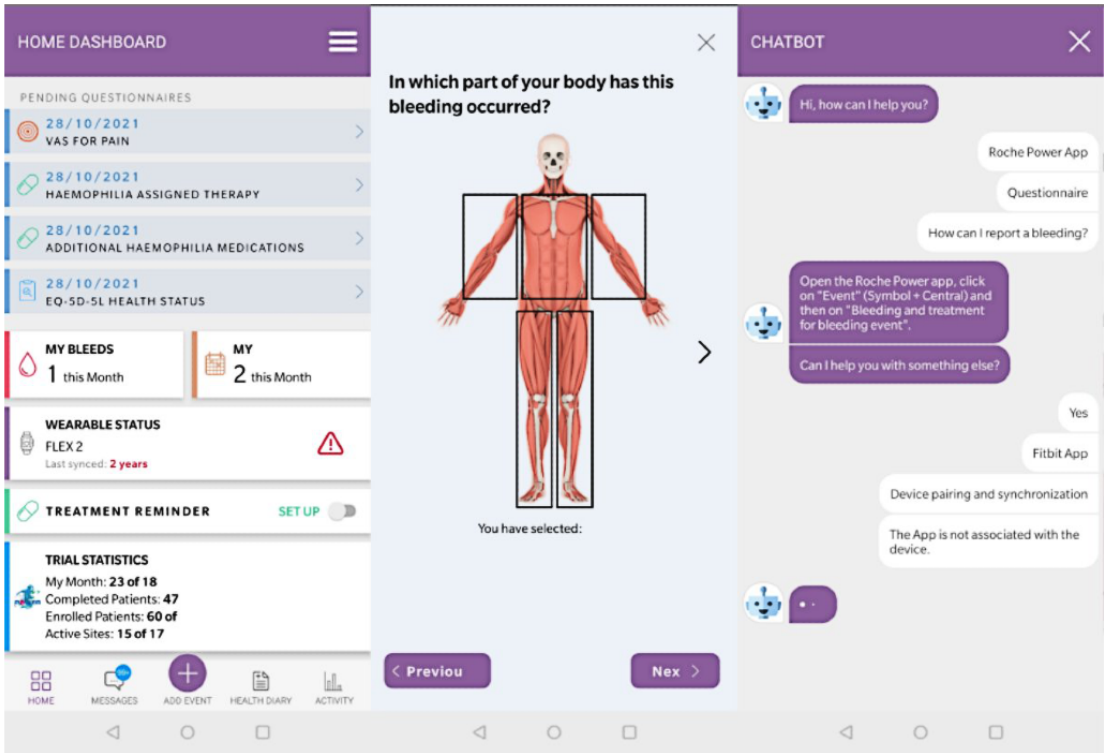
Regarding the *interface design*, the participants only suggested minor revisions in the layout of some questionnaires, particularly to make the reporting of bleeding events more straightforward. Furthermore, one user recommended substituting the human body where bleeds were indicated to make it less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

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



XSL•FO
RenderX

**Figure 2.** Screenshots of the electronic patient-reported outcome app final prototype.




### Usability Evaluation

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

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

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

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

## Discussion

### Principal Findings

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

In contrast, none of the studies intended to support the design of data collection apps, with the exception of a single article that focused on the development process of an app for conducting population surveys, but it was meant to be used by health care professionals only [40].

The design process of the ePRO app for the POWER study is a novel attempt to adopt participatory approaches (and UCD, more specifically) to support the collection of patient-reported data. Although we did not aim to elicit improved self-management behaviors, maximized technology acceptance is a fundamental prerequisite to increase the continuity in data



XSL•FO
RenderX

entry and the likelihood of a study's success. As the precise aims that the app had to pursue were already explicit in the POWER study protocol (ie, to maximize the likelihood of collecting robust outcome data), the design process did not start from scratch and revolved around the *how* rather than the *what*.

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

To sustain user engagement, several of the included features and strategies aimed at powering the intrinsic motivation of the participants, defined elsewhere as altruistic motivation [41]. This is coherent with the belief that although strategies that rely on extrinsic motivation only (*doing something that can lead to an identifiable outcome*) can be effective in the short term, when individuals are intrinsically motivated (*doing something which is felt inherently enriching*), they tend to achieve better results and guarantee their continued participation in the long run [42]. As a result, increased attention was given to the *Study details* and the *Enrollment update* sections, that intend to make the individuals feel they were part of a community and of a mutual effort, which was certified by the growing number of active sites and study participants enrolled.

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

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

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

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

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

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

## Strengths and Limitations

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

Moreover, this study emphasized that user-centered approaches can be applicable, and possibly even more significant, to the development of digital solutions for populations affected with rare diseases, which, by definition, require unique considerations that may be less well-known compared with those for other chronic disease populations. At the same time, the application of the selected methodology to a rare disease group generated additional challenges linked to the difficulties in selecting and



XSL•FO
RenderX

enrolling the participants. Working with smaller sample sizes in rare study populations may be the only way to study them, particularly for methodologies that require the physical presence of participants [53], but several limitations should be acknowledged. First, in terms of patient selection, although we always targeted 5 participants in each of the FGs, in a couple of cases, one of the intended participants communicated their unavailability at short notice because of hemophilia-related issues. Furthermore, to facilitate the identification of the potential participants, we not only adhered to the inclusion criteria identified for the POWER study, but also invited patients with hemophilia A with different severity and inhibitor levels. Although some of the participants were not within the target population of the study, all patients with hemophilia A were subject to similar outcome data collection in standard clinical practice and were, thus, entitled to bring their contribution to the development of the app.

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

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

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

## Conclusions

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

## Acknowledgments

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

## Conflicts of Interest

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

## References

1. Paschou SA, Anagnostis P, Karras S, Annweiler C, Vakalopoulou S, Garipidou V, et al. Bone mineral density in men and children with haemophilia A and B: a systematic review and meta-analysis. Osteoporos Int 2014 Oct;25(10):2399-2407. [doi: 10.1007/s00198-014-2773-7] [Medline: 25001982]
2. Fearn M, Hill K, Williams S, Mudge L, Walsh C, McCarthy P, et al. Balance dysfunction in adults with haemophilia. Haemophilia 2010 Jul 01;16(4):606-614. [doi: 10.1111/j.1365-2516.2010.02200.x] [Medline: 20331756]
3. González LM, Peiró-Velert C, Devís-Devís J, Valencia-Peris A, Pérez-Gimeno E, Pérez-Alenda S, et al. Comparison of physical activity and sedentary behaviours between young haemophilia A patients and healthy adolescents. Haemophilia 2011 Jul;17(4):676-682. [doi: 10.1111/j.1365-2516.2010.02469.x] [Medline: 21299746]
4. Broderick CR, Herbert RD, Latimer J, van Doorn N. Patterns of physical activity in children with haemophilia. Haemophilia 2013 Jan;19(1):59-64. [doi: 10.1111/j.1365-2516.2012.02904.x] [Medline: 22762509]
5. Bouskill V, Hilliard P, Stephens S, Zhang C, Whitney K, Carcao M. An institutional pilot study to investigate physical activity patterns in boys with haemophilia. Haemophilia 2016 Sep;22(5):383-389. [doi: 10.1111/hae.13021] [Medline: 27530715]



XSL•FO
RenderX

6. Sherlock E, O'Donnell JS, White B, Blake C. Physical activity levels and participation in sport in Irish people with haemophilia. Haemophilia 2010 Jan;16(1):202-209. [doi: 10.1111/j.1365-2516.2009.02111.x] [Medline: 19878332]
7. Pérez-Alenda S, Carrasco JJ, Megías-Vericat JE, Poveda JL, Bonanad S, Querol F. Quantification of physical activity in adult patients with haemophilic arthropathy in prophylaxis treatment using a fitness tracker. Haemophilia 2018 Jan;24(1):28-32. [doi: 10.1111/hae.13388] [Medline: 29271553]
8. Versloot O, Berntorp E, Petrini P, Ljung R, Astermark J, Holmström M, et al. Sports participation and physical activity in adult Dutch and Swedish patients with severe haemophilia: A comparison between intermediate- and high-dose prophylaxis. Haemophilia 2019 Mar;25(2):244-251 [FREE Full text] [doi: 10.1111/hae.13683] [Medline: 30690833]
9. Taylor S, Room J, Barker K. Physical activity levels in men with Haemophilia - a single centre UK survey. Haemophilia 2020 Jul;26(4):718-725. [doi: 10.1111/hae.14009] [Medline: 32364278]
10. Baumgardner J, Elon L, Antun A, Stein S, Ribeiro M, Slovensky L, et al. Physical activity and functional abilities in adult males with haemophilia: a cross-sectional survey from a single US haemophilia treatment centre. Haemophilia 2013 Jul;19(4):551-557. [doi: 10.1111/hae.12134] [Medline: 23574421]
11. Allaert FA, Legrand L, Carime NA, Quantin C. Will applications on smartphones allow a generalization of telemedicine? BMC Med Inform Decis Mak 2020 Feb 11;20(1):30 [FREE Full text] [doi: 10.1186/s12911-020-1036-0] [Medline: 32046699]
12. Ciani O, Federici CB. Value lies in the eye of the patients: the why, what, and how of patient-reported outcomes measures. Clin Ther 2020 Jan;42(1):25-33. [doi: 10.1016/j.clinthera.2019.11.016] [Medline: 31932079]
13. Qian W, Lam TT, Lam HH, Li C, Cheung YT. Telehealth interventions for improving self-management in patients with hemophilia: scoping review of clinical studies. J Med Internet Res 2019 Jul 10;21(7):e12340 [FREE Full text] [doi: 10.2196/12340] [Medline: 31293241]
14. Khair K, Holland M. Managing hemophilia: the role of mobile technology. Smart Homecare Technol TeleHealth 2014 May:39-44 [FREE Full text] [doi: 10.2147/shtt.s40961]
15. Boccalandro EA, Dallari G, Mannucci PM. Telemedicine and telerehabilitation: current and forthcoming applications in haemophilia. Blood Transfus 2019 Sep;17(5):385-390 [FREE Full text] [doi: 10.2450/2019.0218-18] [Medline: 30747703]
16. Geraghty S, Dunkley T, Harrington C, Lindvall K, Maahs J, Sek J. Practice patterns in haemophilia A therapy - global progress towards optimal care. Haemophilia 2006 Jan;12(1):75-81. [doi: 10.1111/j.1365-2516.2006.01189.x] [Medline: 16409179]
17. Tarricone R, Petracca F, Ciani O, Cucciniello M. Distinguishing features in the assessment of mHealth apps. Expert Rev Pharmacoecon Outcomes Res 2021 Aug;21(4):521-526. [doi: 10.1080/14737167.2021.1891883] [Medline: 33663324]
18. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
19. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR Mhealth Uhealth 2020 Nov 06;8(11):e22212 [FREE Full text] [doi: 10.2196/22212] [Medline: 32975198]
20. Aiyegbusi OL. Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems. Qual Life Res 2020 Feb;29(2):325-333 [FREE Full text] [doi: 10.1007/s11136-019-02329-z] [Medline: 31691202]
21. Petrini P, Rylander C. Clinical safety surveillance study of the safety and efficacy of long-term home treatment with ReFacto utilizing a computer-aided diary: a Nordic multicentre study. Haemophilia 2009 Jan;15(1):175-183. [doi: 10.1111/j.1365-2516.2008.01863.x] [Medline: 18752534]
22. Hay CR, Xiang H, Scott M, Collins PW, Liesner R, Dolan G, et al. The haemtrack home therapy reporting system: design, implementation, strengths and weaknesses: a report from UK Haemophilia Centre Doctors Organisation. Haemophilia 2017 Sep;23(5):728-735. [doi: 10.1111/hae.13287] [Medline: 28806858]
23. Broderick CR, Herbert RD, Latimer J, Mathieu E, van Doorn N, Curtin JA. Feasibility of short message service to document bleeding episodes in children with haemophilia. Haemophilia 2012 Nov;18(6):906-910. [doi: 10.1111/j.1365-2516.2012.02869.x] [Medline: 22681182]
24. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mHealth. J Health Commun 2012;17 Suppl 1:5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]
25. Huckvale K, Car J. Implementation of mobile health tools. J Am Med Assoc 2014 Apr 09;311(14):1447-1448. [doi: 10.1001/jama.2014.1106] [Medline: 24715083]
26. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
27. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
28. Tremblay MC, Hevner AR, Berndt DJ. Focus groups for artifact refinement and evaluation in design research. Commun Asso Inform Syst 2010;26:Article 27. [doi: 10.17705/1cais.02627]




XSL•FO
RenderX

29. Krueger RA, Casey MA. Focus Groups: A Practical Guide for Applied Research. Thousand Oaks, CA: SAGE Publications Ltd; 2014:1-280.
30. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
31. World Health Organization. Monitoring and Evaluating Digital Health Interventions: A Practical Guide to Conducting Research and Assessment. Geneva, Switzerland: World Health Organization; 2016:1-133.
32. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
33. Jibb LA, Cafazzo JA, Nathan PC, Seto E, Stevens BJ, Nguyen C, et al. Development of a mHealth real-time pain self-management app for adolescents with cancer: an iterative usability testing study [Formula: see text]. J Pediatr Oncol Nurs 2017;34(4):283-294. [doi: 10.1177/1043454217697022] [Medline: 28376666]
34. Lewis JR. Sample sizes for usability studies: additional considerations. Hum Factors 1994 Jun;36(2):368-378. [doi: 10.1177/001872089403600215] [Medline: 8070799]
35. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76. [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
36. Silver C, Lewins A. Using Software in Qualitative Research : A Step-by-Step Guide. Thousand Oaks, CA: SAGE Publications; 2014:1-384.
37. Varshney U. A model for improving quality of decisions in mobile health. Decis Supp Syst 2014 Jun;62:66-77. [doi: 10.1016/j.dss.2014.03.005]
38. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
39. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
40. Shishido HY, da Cruz de Andrade RA, Eler GJ. mHealth data collector: an application to collect and report indicators for assessment of cardiometabolic risk. Stud Health Technol Inform 2014;201:425-432. [Medline: 24943577]
41. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a mobile HIV prevention app for men who have sex with men. JMIR Mhealth Uhealth 2014 Oct 29;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
42. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
43. Tran C, Dicker A, Leiby B, Gressen E, Williams N, Jim H. Utilizing digital health to collect electronic patient-reported outcomes in prostate cancer: single-arm pilot trial. J Med Internet Res 2020 Mar 25;22(3):e12689 [FREE Full text] [doi: 10.2196/12689] [Medline: 32209536]
44. Milward J, Deluca P, Drummond C, Watson R, Dunne J, Kimergård A. Usability testing of the BRANCH smartphone app designed to reduce harmful drinking in young adults. JMIR Mhealth Uhealth 2017 Aug 08;5(8):e109 [FREE Full text] [doi: 10.2196/mhealth.7836] [Medline: 28790022]
45. Giunti G, Kool J, Romero O, Zubiete E. Exploring the specific needs of persons with multiple sclerosis for mhealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
46. Jibb LA, Stevens BJ, Nathan PC, Seto E, Cafazzo JA, Stinson JN. A smartphone-based pain management app for adolescents with cancer: establishing system requirements and a pain care algorithm based on literature review, interviews, and consensus. JMIR Res Protoc 2014 Mar 19;3(1):e15 [FREE Full text] [doi: 10.2196/resprot.3041] [Medline: 24646454]
47. Hartzler AL, BlueSpruce J, Catz SL, McClure JB. Prioritizing the mHealth design space: a mixed-methods analysis of smokers' perspectives. JMIR Mhealth Uhealth 2016 Aug 05;4(3):e95 [FREE Full text] [doi: 10.2196/mhealth.5742] [Medline: 27496593]
48. Pugliese L, Woodriff M, Crowley O, Lam V, Sohn J, Bradley S. Feasibility of the "Bring your own device" model in clinical research: results from a randomized controlled pilot study of a mobile patient engagement tool. Cureus 2016 Mar 16;8(3):e535 [FREE Full text] [doi: 10.7759/cureus.535] [Medline: 27096135]
49. Bromwich M, Bromwich R. Privacy risks when using mobile devices in health care. Can Med Asso J 2016 Sep 06;188(12):855-856 [FREE Full text] [doi: 10.1503/cmaj.160026] [Medline: 27221271]
50. Coons SJ, Eremenco S, Lundy JJ, O'Donohoe P, O'Gorman H, Malizia W. Capturing Patient-Reported Outcome (PRO) data electronically: the past, present, and promise of ePRO measurement in clinical trials. Patient 2015 Aug;8(4):301-309 [FREE Full text] [doi: 10.1007/s40271-014-0090-z] [Medline: 25300613]
51. Ploderer B, Brown R, Seng LS, Lazzarini PA, van Netten JJ. Promoting self-care of diabetic foot ulcers through a mobile phone app: user-centered design and evaluation. JMIR Diabetes 2018 Oct 10;3(4):e10105 [FREE Full text] [doi: 10.2196/10105] [Medline: 30305266]




XSL•FO
RenderX

52. Baek H, Suh J, Kang S, Kang S, Lim TH, Hwang H, et al. Enhancing User Experience Through User Study: Design of an mHealth Tool for Self-Management and Care Engagement of Cardiovascular Disease Patients. JMIR Cardio 2018 Feb 09;2(1):e3. [doi: 10.2196/cardio.9000]
53. Uhlenbusch N, Löwe B, Depping MK. Perceived burden in dealing with different rare diseases: a qualitative focus group study. BMJ Open 2019 Dec 29;9(12):e033353 [FREE Full text] [doi: 10.1136/bmjopen-2019-033353] [Medline: 31888936]
54. Adu MD, Malabu UH, Callander EJ, Malau-Aduli AE, Malau-Aduli BS. Considerations for the development of mobile phone apps to support diabetes self-management: systematic review. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10115 [FREE Full text] [doi: 10.2196/10115] [Medline: 29929949]
55. van Berkel N, Ferreira D, Kostakos V. The experience sampling method on mobile devices. ACM Comput Surv 2018 Jan 12;50(6):1-40. [doi: 10.1145/3123988]

## Abbreviations

**CFG:** confirmatory focus group
**EFG:** exploratory focus group
**ePRO:** electronic patient-reported outcome
**FG:** focus group
**MAUQ:** mHealth app usability questionnaire
**mHealth:** mobile health
**PROM:** patient-reported outcome measure
**UCD:** user-centered design

*Edited by G Eysenbach; submitted 16.10.20; peer-reviewed by E Boccalandro, L Tizek, MDG Pimentel; comments to author 07.12.20; revised version received 28.02.21; accepted 16.09.21; published 01.12.21.*

*Please cite as:*
*Petracca F, Tempre R, Cucciniello M, Ciani O, Pompeo E, Sannino L, Lovato V, Castaman G, Ghirardini A, Tarricone R*
*An Electronic Patient-Reported Outcome Mobile App for Data Collection in Type A Hemophilia: Design and Usability Study*
*JMIR Form Res 2021;5(12):e25071*
*URL: https://formative.jmir.org/2021/12/e25071*
*doi:10.2196/25071*
*PMID:34855619*

©Francesco Petracca, Rosaria Tempre, Maria Cucciniello, Oriana Ciani, Elena Pompeo, Luigi Sannino, Valeria Lovato, Giancarlo Castaman, Alessandra Ghirardini, Rosanna Tarricone. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Use and Appreciation of a Web-Based, Computer-Tailored Diet and Physical Activity Intervention Based on the Self-determination Theory: Evaluation Study of Process and Predictors

Juul M J Coumans[1], MSc; Anke Oenema[2,3], PhD; Catherine A W Bolman[1], PhD; Lilian Lechner[1], PhD

[1]Department of Health Psychology, Faculty of Psychology, Open University of the Netherlands, Heerlen, Netherlands

[2]Department of Health Promotion, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[3]Care and Public Health Research Institute, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

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

## Abstract

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

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

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

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

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

***(JMIR Form Res 2021;5(12):e22390)*** doi:10.2196/22390



XSL•FO
RenderX

**KEYWORDS**

diet; physical activity; eHealth; self-determination theory; motivational interviewing; process evaluation; nonusage attrition

# Introduction

## Background

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

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

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

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

## Objectives

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

# Methods

## Study Design

A 2-group randomized controlled trial (RCT) was conducted in the Netherlands. For this study, observational data of the intervention group of this RCT, called *MyLifestyleCoach*, was used. Therefore, the control group data were excluded. *MyLifestyleCoach* is a web-based, computer-tailored intervention that consists of a diet module (*I Eat*) to promote dietary behavior and a previously tested PA module (*I Move*) to improve PA levels in Dutch adults. Participants in this intervention could choose which of these modules they would like to take part in both modules, the diet module only, the PA



XSL•FO
RenderX

module only, or no module. Detailed information about the development of the intervention and the design of the RCT, which this study is part of, can be found elsewhere [23,24]. This intervention is theoretically founded on the principles of SDT and uses practical applications of motivational interviewing. This intervention was developed using the intervention mapping protocol [25]. This study was reviewed and approved by the Committee for Ethics and Consent in Research of the Open University of the Netherlands (reference U2018/07266/SVW). This study was registered in the Dutch trial register (NL7333). A data processing agreement with the software developer, who acts in line with the General Data Protection Regulation, has been signed. Furthermore, data that have been exported from the software application are safely stored at the servers of the Open University in accordance with the General Data Protection Regulation.

## Participants

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

## Procedure of the Intervention

### *Recruitment*

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

### *Preliminary Assessment and Baseline Questionnaire*

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

### *Opening Session*

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

### *Sessions Within Modules*

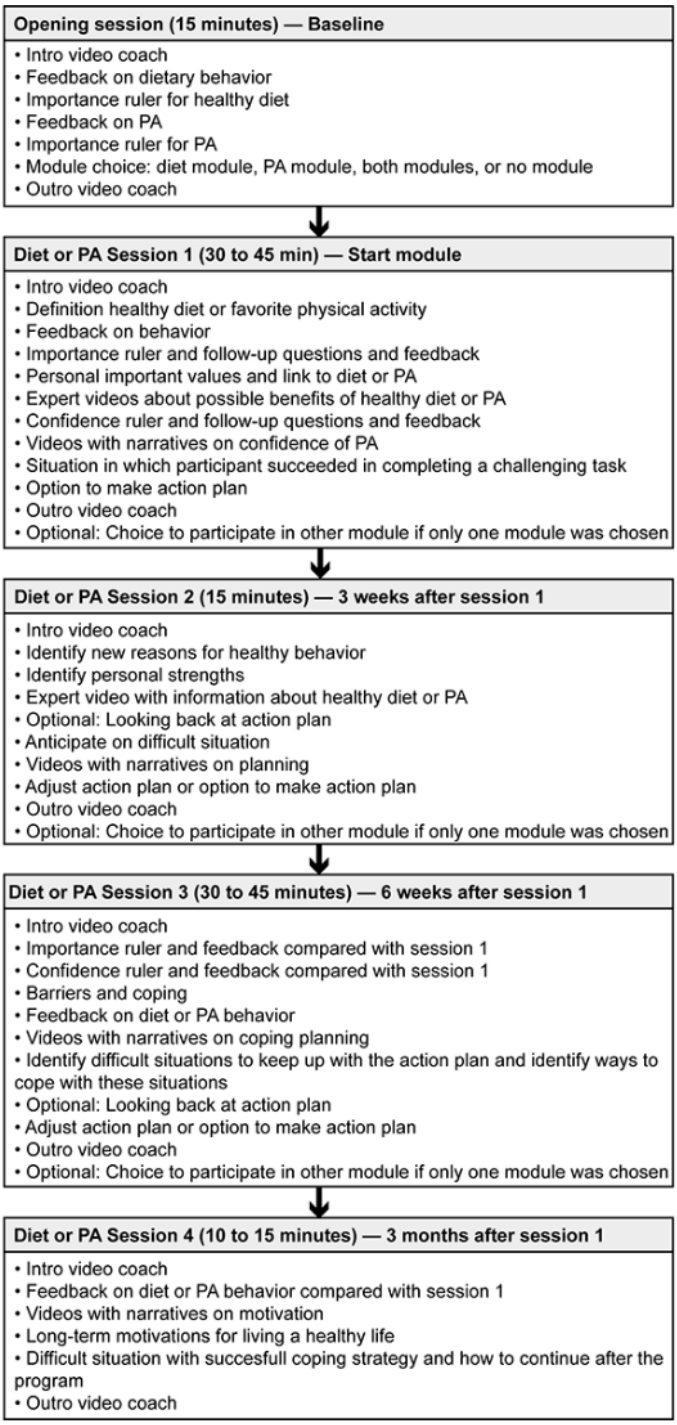
Both the diet and PA modules comprised 4 sessions. In *session 1*, a healthy diet was explained according to the Netherlands Nutrition Center, or guidelines for sufficient PA levels were provided. Participants were able to see their results on their dietary or PA behavior again. The importance and confidence in eating (more) healthily or engaging in sufficient PA levels were assessed, and feedback was given on this topic. Finally, participants could make an action plan. After 3 weeks from the first session, participants could enter *session 2*. In this session, they looked back on their perception of the importance of a healthy diet or PA level. They could come up with new reasons to start with the new behavior. Furthermore, they thought about what effects it would have on them if they started with the new behavior (*looking forward*). Finally, they engaged in a part on coping with difficult situations, including the identification of personal strengths, and could formulate or change the action plan. After 6 weeks from session 1, participants filled in a short questionnaire about their current behavior (diet or PA) and then entered *session 3*. An assessment took place on participants' current perception of the importance of a healthy diet or PA and their confidence in achieving this behavior compared with session 1. Participants also received feedback on this assessment. Participants were invited to think back on a problematic situation in which they struggled but managed the achieved behavior.



XSL•FO
RenderX

They received feedback on their current dietary or PA behavior compared with session 1 and could formulate or change the action plan expressed in the previous session of sessions. After 3 months from session 1, they filled in a short questionnaire regarding their current behavior (diet or PA) and then entered *session 4*, which served as a booster session. Participants could choose several topics from previous sessions that they wanted to do. These topics included feedback on their current behavior compared with session 1, long-term personal motivation and confidence, how to deal with difficult situations, and information on how to maintain their new behavior after the end of the program. Figure 1 shows an overview of the content of these sessions. More detailed information on these sessions can be found in the protocol papers [23,24].

**Figure 1.** Overview of the content of the sessions in the intervention. PA: physical activity.






XSL•FO
RenderX

## Follow-up Questionnaire

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

## Measurements

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

### *Demographics*

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

### *Motivation*

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

### *Dietary and PA Behavior*

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

## Process Evaluation

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

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

## Statistical Analysis

Descriptive statistics (mean and SD values) and frequencies (and percentages) were used to depict the characteristics of the participants, the overall flow through the intervention, and appreciation scores. Logistic regression analyses were conducted to examine which personal characteristics (age, gender, education, marital status, work, physical impairment, health status, and BMI) and motivational characteristics were related to use. Use was subdivided into 3 parts according to different points in the intervention: (1) completion of the opening session; (2) the initial module's session, that is, completion of the first session of the diet module when only the diet module was chosen or the first session of the diet and PA module when both modules were chosen; and (3) the follow-up sessions, that is, completion of all 4 sessions of the diet module when only the diet module was chosen or completion of all 8 sessions of the diet and PA module when both modules were chosen. Furthermore, linear regression analyses were performed to investigate which demographic factors, motivational constructs, and program features were associated with the intervention appreciation scores (overall intervention, diet, and PA module).



XSL•FO
RenderX

All statistical analyses were performed using the statistical software R (version 3.6.0; R Foundation for Statistical Computing). For all regression analyses, a stepwise approach was used in which the demographic variables were entered in the first step, motivational constructs were introduced in the second step, and program features were added in the third step. Variance inflation factors were inspected before conducting the analyses. Statistical significance was set at $P<.05$.

## Results

### Participants' Characteristics

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



XSL•FO
RenderX

**Table 1.** Demographic characteristics of the full sample (N=775).

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

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

[b]PA: physical activity.

[c]MVPA: moderate to vigorous physical activity.



XSL•FO
RenderX

### Description of the Participants' Flow and Module Use

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

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

Within the PA module (n=12, as 2 participants did not receive an invitation mail for the first session), half of the participants decided to start immediately, and the other half wanted to start later. Of the 6 participants who decided to start immediately, 4 (67%) completed session 1, 2 (33%) completed session 2, 1 (17%) completed session 3, and 1 (17%) completed the fourth session. Of the 6 participants who decided to start later, only 1 (17%) completed all 4 sessions. Participants who only chose the diet module were asked after every session, except for the last one, whether they were interested in starting with the PA module. Of the 17 participants who chose to start with the PA module, 8 (47%) completed the first session of the PA module, 7 (41%) completed the second and third sessions, and 4 (24%) completed the last session. Participants who only chose the PA module were asked after every session, except for the last one, whether they were interested in starting with the diet module. Of the 12 participants who only chose the PA module, 1 (8%) participant was interested in the diet module; however, this person did not complete any session of the diet module.

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

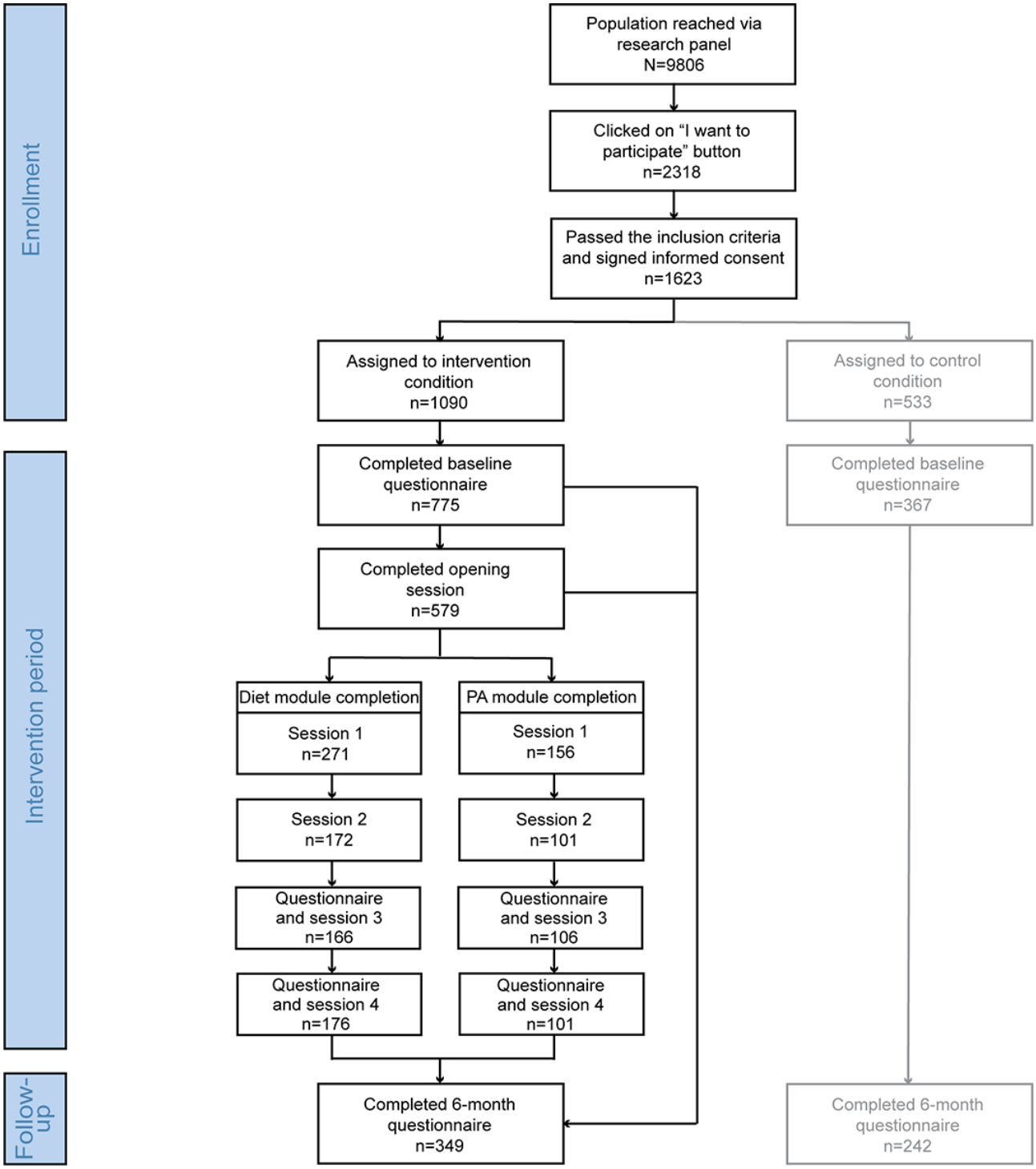
Participants who initially chose no module in the intervention (110/775, 14.2%) were sent an email directing them to the website where they could change their choice. Of the 110 participants, only 4 (3.6%) reconsidered their choice, and 2 (1.8%) chose a module. Of the 110 participants, 1 (0.9%) chose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

Finally, 45% (349/775) of participants completed the follow-up questionnaire at 6 months from baseline. Of the 158 participants who chose the diet module, 78 (49.4%) completed the follow-up questionnaire. Of the 14 participants who chose the PA module, 9 (64.3%) completed the follow-up questionnaire. Of the 339 participants who chose both modules, 151 (44.5%) completed the follow-up questionnaire. Of the 108 participants who did not choose any module in the opening session, 58 (53.7%) completed the follow-up questionnaire. Of the 156 participants who did not enter the opening session or made no module choice, 53 (33.9%) completed the follow-up questionnaire. This follow-up measurement included the process evaluation questions.



XSL•FO
RenderX

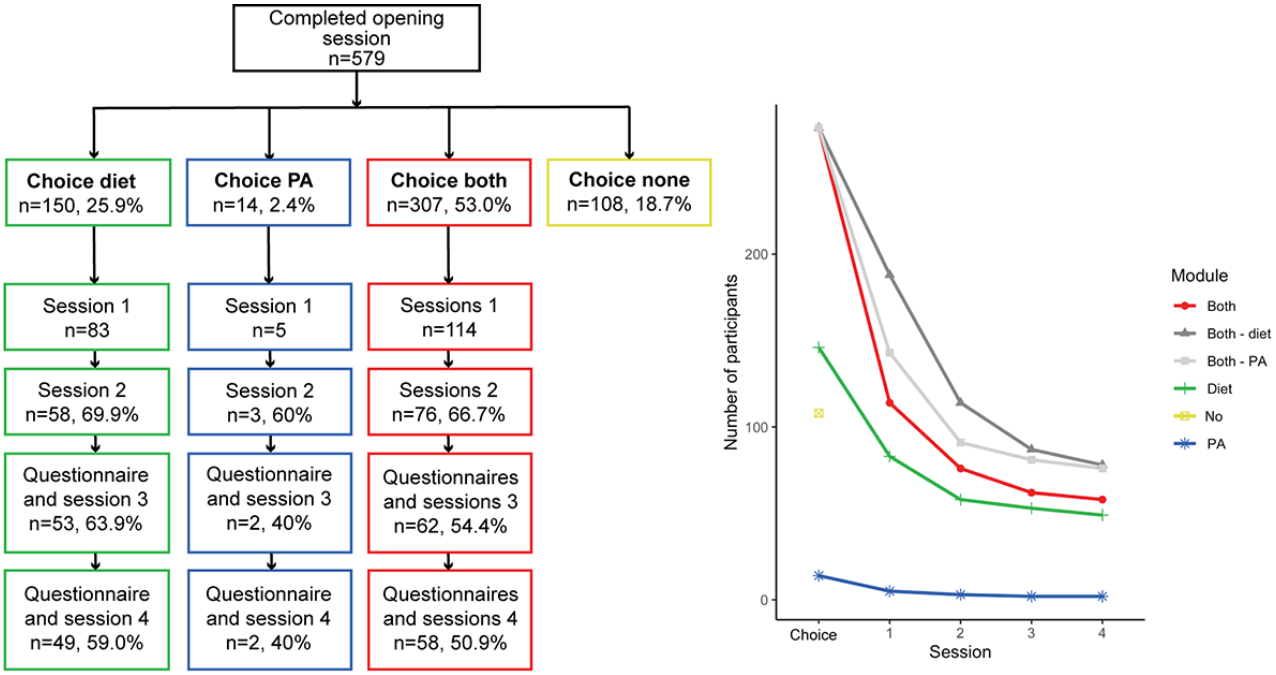
**Figure 2.** Participants' flow in the randomized controlled trial. Participants did not need to complete the second session to be able to continue the third session. Therefore, use rates do not necessarily represent a funnel shape. PA: physical activity.






XSL•FO
RenderX

**Figure 3.** Completion rates (left panel) and curves (right panel) stratified for module choice. PA: physical activity.




## Predictors of Use of Initial and Follow-up Sessions

As described in the previous section and as can be seen in the participants' flowchart (Figure 2) and the completion rates in Figure 3, there are several moments within the intervention at which participants stop using the intervention. First, several people did not complete the baseline questionnaire after randomization (315/1090, 28.9%; Figure 2). As we did not have the demographic characteristics of this group, it was not possible to further examine predictors of why they stopped using the intervention. Second, there was a significant number of participants who did not complete the opening session after completing the baseline questionnaire (Figure 2). Third, another group did not start or end the first session of their chosen module (Figure 3). Fourth, numerous people did not complete the sessions as intended, that is, about half of the participants completed the whole intervention once started (Figure 3). Here, we investigate whether there are characteristics associated with use for the latter 3 moments.

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

The logistic regression analyses relating demographic characteristics, motivational constructs, and intervention features with completing the first session showed that people who had a physical impairment and those with higher scores on autonomous motivation to eat healthily were more likely to complete the first session of the diet module. On the other hand, participants with a higher BMI or more intrinsic motivation to eat healthily were less likely to complete the first session of the diet module. For participants who chose both modules, the results showed that participants having a higher self-reported health status, more controlled motivation to PA, and receiving red advice for PA compared with orange advice were more likely to complete the first sessions of both modules. On the contrary, participants with a partner or those who preferred to start with the PA module were less likely to complete the first sessions of both modules. The results are presented in Table 3. Owing to the low number of participants that only chose the PA module, predictors of use for the PA module were not further investigated, as it would have been statistically invalid.

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



XSL•FO
RenderX

**Table 2.** Results of the stepwise logistic regression analyses (full model) showing variables associated with completing the opening session (N=775)[a].

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

[a]The results' interpretations are reported when all other predictors are held constant. Explained variance $R^2$ tjur=0.054; Akaike information criterion=869.59.
[b]*Female* is the reference category.
[c]*Medium education* is the reference category.
[d]*Single* is the reference category.
[e]Values represent statistical significance.
[f]*Being unemployed* is the reference category.
[g]*No physical impairment* is the reference category.
[h]PA: physical activity.



XSL•FO
RenderX

**Table 3.** Results of the stepwise logistic regression analyses (full model) showing variables associated with completing the first session for the diet module and both modules.

| Predictors | Session 1 (1=yes and 0=no) | | | |
|---|---|---|---|---|
| | Diet module[a,b] | | Both modules[c] | |
| | OR[d] (95% CI; SE) | *P* value | OR (95% CI; SE) | *P* value |
| Intercept | 2.88 (0.04-221.08; 2.22) | .63 | 0.06 (0.002-2.39; 1.85) | .14 |
| Age | 1.02 (0.99-1.06; 0.02) | .26 | 1.02 (0.99-1.04; 0.01) | .15 |
| Gender[e] | 0.69 (0.31-1.56; 0.41) | .38 | 0.70 (0.38-1.31; 0.32) | .27 |
| Education high[f] | 1.02 (0.38-2.73; 0.50) | .96 | 1.39 (0.73-2.66; 0.33) | .32 |
| Education low[f] | 1.53 (0.19-12.44; 1.07) | .69 | 1.21 (0.25-5.91; 0.81) | .82 |
| Marital status partner[g] | 1.03 (0.43-2.45; 0.44) | .95 | 0.52 (0.28-0.95; 0.31) | .03[h] |
| Work employed[i] | 1.26 (0.53-3.01; 0.44) | .60 | 0.84 (0.46-1.54; 0.31) | .57 |
| Impairment[j] | 9.05 (1.06-77.10; 1.09) | .04[h] | 3.04 (0.79-11.75; 0.69) | .12 |
| BMI | 0.87 (0.79-0.97; 0.05) | .009[h] | 0.98 (0.93-1.04; 0.03) | .55 |
| Health status | 1.02 (0.99-1.05; 0.02) | .25 | 1.04 (1.01-1.06; 0.01) | .001[h] |
| Amotivation diet | 0.83 (0.49-1.39; 0.26) | .47 | 1.19 (0.84-1.67; 0.17) | .33 |
| Amotivation PA[k] | 1.50 (0.95-2.37; 0.23) | .08 | 0.72 (0.49-1.06; 0.20) | .10 |
| Controlled motivation diet | 1.24 (0.68-2.25; 0.30) | .48 | 0.70 (0.47-1.06; 0.21) | .09 |
| Controlled motivation PA | 0.65 (0.37-1.15; 0.29) | .14 | 1.70 (1.13-2.57; 0.21) | .01[h] |
| Autonomous motivation diet | 2.27 (1.17-4.41; 0.34) | .02[h] | 0.995 (0.64-1.55; 0.23) | .98 |
| Autonomous motivation PA | 0.69 (0.36-1.33; 0.33) | .27 | 0.97 (0.61-1.55; 0.24) | .90 |
| Intrinsic motivation diet | 0.53 (0.31-0.91; 0.27) | .02[h] | 0.94 (0.66-1.33; 0.18) | .71 |
| Intrinsic motivation PA | 1.05 (0.65-1.33; 0.25) | .83 | 0.80 (0.58-1.11; 0.17) | .19 |
| Diet advice green[l,m] | —[n] | — | 0.00 (974.90) | .99 |
| Diet advice red[l] | 2.77 (0.98-7.86; 0.53) | .06 | 0.59 (0.31-1.12; 0.33) | .11 |
| PA advice green[l] | — | — | 3.18 (0.40-24.92; 1.05) | .27 |
| PA advice red[l] | — | — | 16.82 (1.28-221.17; 1.31) | .03[h] |
| Module start (later[o]) | 0.55 (0.25-1.22; 0.40) | .14 | — | — |
| First module (PA[p]) | — | — | 0.21 (0.07-0.63; 0.56) | .005[h] |

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

[b]Observations=146; $R^2$ tjur=0.158; Akaike information criterion=215.19.

[c]Observations=273; $R^2$ tjur=0.167; Akaike information criterion=366.99.

[d]OR: odds ratio.

[e]*Female* is the reference category.

[f]*Medium* education is the reference category.

[g]*Single* is the reference category.

[h]Values represent statistical significance.

[i]*Being* unemployed is the reference category.

[j]*No physical impairment* is the reference category.

[k]PA: physical activity.

[l]*Orange advice* is the reference category.

[m]Only 2 participants received green advice. Consequently, the odds ratio and SE are less reliable, and CI is not reported.



XSL•FO
RenderX

[n]These variables were not included in the model.

[o]*Directly starting with the first session* was the reference category.

[p]*Choosing the diet module to start with when both modules were chosen* was the reference category.



XSL•FO
RenderX

**Table 4.** Results of the stepwise logistic regression analyses (full model) showing variables associated with completing all sessions when the diet or both modules were chosen[a].

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

[a]The results' interpretations are reported when all other predictors are held constant.
[b]Observations=146; $R^2$ tjur=0.099; Akaike information criterion=211.14.
[c]Observations=273; $R^2$ tjur=0.188; Akaike information criterion=276.69.
[d]OR: odds ratio.
[e]Values represent statistical significance.
[f]*Female* is the reference category.
[g]*Medium education* is the reference category.
[h]*Single* is the reference category.
[i]*Being unemployed* is the reference category.
[j]*No physical impairment* is the reference category.
[k]PA: physical activity.
[l]*Orange advice* is the reference category.
[m]Only 2 participants received green advice. Consequently, the odds ratio and SE are less reliable, and CI is not reported.
[n]These variables were not included in the model.



XSL•FO
RenderX

[o]*Directly starting with the first session* was the reference category.

[p]*Choosing the diet module to start with when both modules were chosen* was the reference category.

## Appreciation and Its Predictors

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

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



XSL•FO
RenderX

**Table 5.** Results of the stepwise regression analyses (full model) showing variables associated with appreciation scores[a].

| Predictors | Appreciation[b] | | | Appreciation of diet module[c] | | | Appreciation of PA[d] module[e] | | |
|---|---|---|---|---|---|---|---|---|---|
| | b[f] (SE) | B[g] | *P* value | b (SE) | B | *P* value | b (SE) | B | *P* value |
| Intercept | 0.43 (1.17) | 0.00 | .71 | 1.17 (1.35) | 0.00 | .39 | 0.36 (1.63) | 0.00 | .83 |
| Age (years) | 0.01 (0.01) | 0.06 | .35 | −0.00 (0.01) | −0.03 | .71 | −0.01 (0.01) | −0.05 | .57 |
| Gender[h] | −0.11 (0.18) | −0.03 | .56 | −0.15 (0.20) | −0.04 | .45 | 0.19 (0.27) | 0.05 | .49 |
| Education high[i] | 0.14 (0.21) | 0.04 | .50 | 0.05 (0.22) | 0.01 | .81 | −0.05 (0.28) | −0.01 | .85 |
| Education low[i] | 0.28 (0.46) | 0.03 | .54 | −0.25 (0.43) | −0.03 | .57 | 0.14 (0.71) | 0.01 | .85 |
| Marital partner[j] | 0.21 (0.19) | 0.06 | .29 | 0.03 (0.21) | 0.01 | .88 | 0.04 (0.29) | 0.01 | .89 |
| Work[k] | 0.0003 (0.19) | 0.0001 | .99 | 0.07 (0.20) | 0.02 | .71 | 0.05 (0.28) | 0.01 | .86 |
| Impairment[l] | −0.73 (0.41) | −0.10 | .08 | −0.09 (0.39) | −0.01 | .82 | −0.17 (0.49) | −0.03 | .72 |
| BMI | 0.01 (0.02) | 0.03 | .55 | −0.01 (0.02) | −0.04 | .59 | −0.02 (0.02) | −0.09 | .26 |
| Health status | 0.006 (0.01) | 0.04 | .52 | −0.002 (0.01) | −0.02 | .80 | −0.01 (0.01) | −0.05 | .56 |
| Amotivation diet | 0.14 (0.11) | 0.10 | .20 | −0.10 (0.11) | −0.07 | .38 | −0.37 (0.18) | −0.24 | .05[n] |
| Amotivation PA | 0.0003 (0.11) | 0.0002 | .99 | 0.24 (0.12) | 0.17 | .05[n] | 0.34 (0.17) | 0.23 | .06 |
| Controlled motivation Diet | 0.07 (0.12) | 0.05 | .56 | −0.09 (0.15) | −0.07 | .53 | 0.02 (0.17) | 0.02 | .90 |
| Controlled motivation PA | −0.14 (0.12) | −0.10 | .27 | −0.09 (0.14) | −0.07 | .51 | −0.15 (0.17) | −0.11 | .41 |
| Autonomous motivation diet | −0.05 (0.14) | −0.03 | .74 | 0.25 (0.19) | 0.17 | .18 | −0.10 (0.19) | −0.07 | .61 |
| Autonomous motivation PA | 0.23 (0.14) | 0.15 | .09 | 0.09 (0.19) | 0.06 | .64 | 0.52 (0.20) | 0.37 | .01[n] |
| Intrinsic motivation diet | 0.19 (0.12) | 0.10 | .11 | 0.31 (0.13) | 0.18 | .02[n] | 0.12 (0.18) | 0.07 | .51 |
| Intrinsic motivation PA | −0.12 (0.10) | −0.08 | .23 | −0.09 (0.11) | −0.06 | .40 | −0.04 (0.15) | −0.03 | .82 |
| Autonomy | 0.22 (0.13) | 0.10 | .10 | 0.17 (0.15) | 0.06 | .27 | 0.32 (0.21) | 0.12 | .13 |
| Relatedness | 0.31 (0.18) | 0.16 | .10 | 0.38 (0.17) | 0.20 | .03[m] | 0.51 (0.25) | 0.26 | .05[m] |
| Competence | 0.59 (0.14) | 0.34 | <.001[m] | 0.78 (0.13) | 0.49 | <.001[m] | 0.52 (0.18) | 0.34 | .005[m] |
| Diet advice green[n] | 1.25 (1.43) | 0.04 | .38 | __[o] | — | — | — | — | — |
| Diet advice red[n] | 0.14 (0.21) | 0.03 | .49 | −0.07 (0.22) | −0.02 | .74 | −0.22 (0.29) | −0.05 | .45 |
| PA advice green[n] | 0.43 (0.61) | 0.06 | .48 | 0.04 (0.68) | 0.01 | .96 | 0.44 (0.77) | 0.08 | .57 |
| PA advice red[n] | −0.32 (0.73) | −0.04 | .66 | −0.26 (0.81) | −0.03 | .74 | 0.54 (0.88) | 0.09 | .54 |
| Module choice diet[p] | −0.45 (0.31) | −0.11 | .14 | −0.26 (0.19) | −0.07 | .18 | −0.14 (0.51) | −0.02 | .79 |
| Module choice PA[p] | 0.32 (0.54) | 0.03 | .55 | — | — | — | 0.29 (0.70) | 0.03 | .68 |
| Module choice both[p] | −0.58 (0.27) | −0.17 | .03[m] | — | — | — | — | — | — |
| Sessions diet | 0.03 (0.07) | 0.04 | .61 | −0.04 (0.10) | −0.02 | .67 | — | — | — |
| Sessions PA | 0.15 (0.07) | 0.15 | .04[m] | — | — | — | 0.27 (0.16) | 0.13 | .09 |

[a]The results' interpretations are reported when all other predictors are held constant.

[b]Observations=291; $R^2/R^2$ adjusted=0.431/0.368; Akaike information criterion=1047.00.

[c]Observations=159; $R^2/R^2$ adjusted=0.669/0.607; Akaike information criterion=494.01.

[d]PA: physical activity.

[e]Observations=101; $R^2/R^2$ adjusted=0.726/0.630; Akaike information criterion=318.10.

[f]b: unstandardized regression coefficient.

[g]B: standardized regression coefficient.



XSL•FO
RenderX

[h]*Female* is the reference category.

[i]*Medium education* is the reference category.

[j]*Single* is the reference category.

[k]*Being unemployed* is the reference category.

[l]*No physical impairment* is the reference category.

[m]Values represent statistical significance.

[n]*Orange advice* is the reference category.

[o]These variables were not included in the model.

[p]*Choosing no module* is the reference category in the general intervention's appreciation, whereas *choosing both modules* is the reference category for the appreciation of the diet module and the physical activity module.

## Discussion

### Principal Findings

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

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

Some program-related features are also linked to use, such as the advice a person received at the start of the intervention based on an assessment of the dietary and PA behavior on his or her initial behavioral performance. For instance, people with red advice for diet, indicating much room for improvement in their dietary behavior, were less likely to complete the whole intervention and thus both modules but were more likely to complete the diet module when only the diet module was chosen. The red diet advice seemed to increase participants' focus to follow the diet sessions while decreasing their broader participation, possibly because of a high load or ego depletion and a lack of mental resources [38]. Therefore, it is important to keep in mind whether it is beneficial to provide advice in an intervention, as this may either have negative or positive effects on use. Another finding was that giving participants a choice when to start with the module was not predictive of completion. Thus, postponing the start of a module does not necessarily lead to nonusage attrition.

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



XSL•FO
RenderX

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

## Implications

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

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

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

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

## Limitations

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

## Conclusions

This process evaluation study adds meaningful information on the use and appreciation of a web-based, computer-tailored dietary and PA intervention—*MyLifestyleCoach*. The results indicate that different types of motivation that were examined in this study at play at other moments where people are likely to stop using the intervention, such as the initial session or sessions or completing the whole intervention. Appreciation was associated with several motivational constructs, such as amotivation and intrinsic motivation, and related to basic psychological needs, such as competence. We derived some practical implications for developing eHealth interventions that contain multiple health behaviors. For instance, we found that



XSL•FO
RenderX

about half of the participants ended the entire intervention once they finished the first session. Therefore, we recommend making the first session in a multiple-session intervention short, challenging, and rewarding and allow the person to choose small goals and achieve success.

## Acknowledgments

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

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Results from the stepwise logistic regression predicting completion of intervention components.
[DOCX File , 53 KB - formative_v5i12e22390_app1.docx ]

Multimedia Appendix 2
Results from the stepwise linear regression predicting the appreciation score for the whole intervention, the diet module, and physical activity module.
[DOCX File , 45 KB - formative_v5i12e22390_app2.docx ]

## References

1. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
2. McVay MA, Bennett GG, Steinberg D, Voils CI. Dose-response research in digital health interventions: concepts, considerations, and challenges. Health Psychol 2019 Dec;38(12):1168-1174 [FREE Full text] [doi: 10.1037/hea0000805] [Medline: 31580127]
3. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
4. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
5. Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. Annu Rev Public Health 2007;28:413-433. [doi: 10.1146/annurev.publhealth.28.021406.144145] [Medline: 17150029]
6. Robroek SJ, Brouwer W, Lindeboom D, Oenema A, Burdorf A. Demographic, behavioral, and psychosocial correlates of using the website component of a worksite physical activity and healthy nutrition promotion program: a longitudinal study. J Med Internet Res 2010 Sep 30;12(3):e44 [FREE Full text] [doi: 10.2196/jmir.1402] [Medline: 20921001]
7. Schneider F, van Osch L, Schulz DN, Kremers SP, de Vries H. The influence of user characteristics and a periodic email prompt on exposure to an internet-delivered computer-tailored lifestyle program. J Med Internet Res 2012 Mar 01;14(2):e40 [FREE Full text] [doi: 10.2196/jmir.1939] [Medline: 22382037]
8. Schulz DN, Schneider F, de Vries H, van Osch LA, van Nierop PW, Kremers SP. Program completion of a web-based tailored lifestyle intervention for adults: differences between a sequential and a simultaneous approach. J Med Internet Res 2012 Mar 08;14(2):e26 [FREE Full text] [doi: 10.2196/jmir.1968] [Medline: 22403770]
9. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov 04;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
10. Van't Riet J, Crutzen R, De Vries H. Investigating predictors of visiting, using, and revisiting an online health-communication program: a longitudinal study. J Med Internet Res 2010 Sep 02;12(3):e37 [FREE Full text] [doi: 10.2196/jmir.1345] [Medline: 20813716]
11. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A self-regulation-based eHealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
12. Brouwer W, Oenema A, Raat H, Crutzen R, de Nooijer J, de Vries NK, et al. Characteristics of visitors and revisitors to an internet-delivered computer-tailored lifestyle intervention implemented for use by the general public. Health Educ Res 2010 Aug;25(4):585-595. [doi: 10.1093/her/cyp063] [Medline: 19897515]
13. Reinwand DA, Schulz DN, Crutzen R, Kremers SP, de Vries H. Who follows eHealth interventions as recommended? A study of participants' personal characteristics from the experimental arm of a randomized controlled trial. J Med Internet Res 2015 May 11;17(5):e115 [FREE Full text] [doi: 10.2196/jmir.3932] [Medline: 25963607]



XSL•FO
RenderX

14. Leung AW, Chan RS, Sea MM, Woo J. An overview of factors associated with adherence to lifestyle modification programs for weight management in adults. Int J Environ Res Public Health 2017 Aug 16;14(8):922 [FREE Full text] [doi: 10.3390/ijerph14080922] [Medline: 28813030]
15. Rothman A, Baldwin A, Hertel A, Fuglestad P. Self-regulation and behavior change: disentangling behavioral initiation and behavioral maintenance. In: Handbook of Self-Regulation: Research, Theory, and Applications. New York, NY: The Guilford Press; 2011.
16. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
17. Coa K, Patrick H. Baseline motivation type as a predictor of dropout in a healthy eating text messaging program. JMIR Mhealth Uhealth 2016 Sep 29;4(3):e114 [FREE Full text] [doi: 10.2196/mhealth.5992] [Medline: 27688034]
18. Ryan RM, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: interventions based on self-determination theory. Eur Health Psychol 2008;10(1):2-5 [FREE Full text]
19. Fortier MS, Duda JL, Guerin E, Teixeira PJ. Promoting physical activity: development and testing of self-determination theory-based interventions. Int J Behav Nutr Phys Act 2012 Mar 02;9:20 [FREE Full text] [doi: 10.1186/1479-5868-9-20] [Medline: 22385751]
20. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
21. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol 2008;49(3):182-185 [FREE Full text] [doi: 10.1037/a0012801]
22. Coumans JM, Bolman CA, Oenema A, Lechner L. Predictors of self-determined module choice in a web-based computer-tailored diet and physical activity intervention: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2020 Jul 23;22(7):e15024 [FREE Full text] [doi: 10.2196/15024] [Medline: 32706659]
23. Friederichs SA, Oenema A, Bolman C, Guyaux J, van Keulen HM, Lechner L. I Move: systematic development of a web-based computer tailored physical activity intervention, based on motivational interviewing and self-determination theory. BMC Public Health 2014 Feb 28;14:212 [FREE Full text] [doi: 10.1186/1471-2458-14-212] [Medline: 24580802]
24. Coumans JM, Bolman CA, Friederichs SA, Oenema A, Lechner L. Development and testing of a personalized web-based diet and physical activity intervention based on motivational interviewing and the self-determination theory: protocol for the MyLifestyleCoach randomized controlled trial. JMIR Res Protoc 2020 Feb 04;9(2):e14491 [FREE Full text] [doi: 10.2196/14491] [Medline: 32014841]
25. Bartholomew LK, Markham CM, Ruiter RA, Fernàndez ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach. Fourth edition. San Francisco, CA: Jossey-Bass Inc; 2016.
26. Brink E, van Rossum C, Postma-Smeets A, Stafleu A, Wolvers D, van Dooren C, et al. Development of healthy and sustainable food-based dietary guidelines for the Netherlands. Public Health Nutr 2019 Sep;22(13):2419-2435 [FREE Full text] [doi: 10.1017/S1368980019001435] [Medline: 31262374]
27. Weggemans RM, Backx FJ, Borghouts L, Chinapaw M, Hopman MT, Koster A, Committee Dutch Physical Activity Guidelines 2017. The 2017 Dutch physical activity guidelines. Int J Behav Nutr Phys Act 2018 Jun 25;15(1):58 [FREE Full text] [doi: 10.1186/s12966-018-0661-9] [Medline: 29940977]
28. Levesque CS, Williams GC, Elliot D, Pickering MA, Bodenhamer B, Finley PJ. Validating the theoretical structure of the Treatment Self-Regulation Questionnaire (TSRQ) across three different health behaviors. Health Educ Res 2007 Oct;22(5):691-702. [doi: 10.1093/her/cyl148] [Medline: 17138613]
29. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. J Sport Exerc Psychol 2004 Jun;26(2):191-196. [doi: 10.1123/jsep.26.2.191]
30. Huybrechts I, Börnhorst C, Pala V, Moreno LA, Barba G, Lissner L, IDEFICS Consortium. Evaluation of the Children's Eating Habits Questionnaire used in the IDEFICS study by relating urinary calcium and potassium to milk consumption frequencies among European children. Int J Obes (Lond) 2011 Apr;35(1):S69-S78. [doi: 10.1038/ijo.2011.37] [Medline: 21483425]
31. Coumans JM, Danner UN, Intemann T, De Decker A, Hadjigeorgiou C, Hunsberger M, I.Family Consortium. Emotion-driven impulsiveness and snack food consumption of European adolescents: results from the I.Family study. Appetite 2018 Apr 01;123:152-159. [doi: 10.1016/j.appet.2017.12.018] [Medline: 29269316]
32. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003 Dec;56(12):1163-1169. [doi: 10.1016/s0895-4356(03)00220-8] [Medline: 14680666]
33. Walthouwer MJ, Oenema A, Lechner L, de Vries H. Comparing a video and text version of a web-based computer-tailored intervention for obesity prevention: a randomized controlled trial. J Med Internet Res 2015 Oct 19;17(10):e236 [FREE Full text] [doi: 10.2196/jmir.4083] [Medline: 26481772]
34. Friederichs SA, Oenema A, Bolman C, Lechner L. Long term effects of self-determination theory and motivational interviewing in a web-based physical activity intervention: randomized controlled trial. Int J Behav Nutr Phys Act 2015 Aug 18;12:101 [FREE Full text] [doi: 10.1186/s12966-015-0262-9] [Medline: 26283094]




XSL•FO
RenderX

35. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
36. Ahnis A, Riedl A, Figura A, Steinhagen-Thiessen E, Liebl ME, Klapp BF. Psychological and sociodemographic predictors of premature discontinuation of a 1-year multimodal outpatient weight-reduction program: an attrition analysis. Patient Prefer Adherence 2012;6:165-177 [FREE Full text] [doi: 10.2147/PPA.S28022] [Medline: 22442628]
37. Vallerand R. Toward a hierarchical model of intrinsic and extrinsic motivation. Adv Exp Soc Psychol 1997;29:271-360. [doi: 10.1016/s0065-2601(08)60019-2]
38. Baumeister RF, Bratslavsky E, Muraven M, Tice DM. Ego depletion: is the active self a limited resource? J Pers Soc Psychol 1998;74(5):1252-1265. [doi: 10.1037/0022-3514.74.5.1252]
39. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
40. Woodall WG, Buller DB, Saba L, Zimmerman D, Waters E, Hines JM, et al. Effect of emailed messages on return use of a nutrition education website and subsequent changes in dietary behavior. J Med Internet Res 2007 Sep 30;9(3):e27 [FREE Full text] [doi: 10.2196/jmir.9.3.e27] [Medline: 17942389]
41. Hebert JR, Hurley TG, Peterson KE, Resnicow K, Thompson FE, Yaroch AL, et al. Social desirability trait influences on self-reported dietary measures among diverse participants in a multicenter multiple risk factor trial. J Nutr 2008 Jan;138(1):226S-234S. [doi: 10.1093/jn/138.1.226S] [Medline: 18156429]
42. Rhodes SD, Bowie DA, Hergenrather KC. Collecting behavioural data using the world wide web: considerations for researchers. J Epidemiol Community Health 2003 Jan;57(1):68-73 [FREE Full text] [doi: 10.1136/jech.57.1.68] [Medline: 12490652]
43. Pampel FC, Krueger PM, Denney JT. Socioeconomic disparities in health behaviors. Annu Rev Sociol 2010 Aug;36:349-370 [FREE Full text] [doi: 10.1146/annurev.soc.012809.102529] [Medline: 21909182]

## Abbreviations

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

*Edited by R Kukafka, G Eysenbach; submitted 10.07.20; peer-reviewed by N Nicolay, S Karnes; comments to author 30.09.20; revised version received 23.11.20; accepted 29.09.21; published 02.12.21.*

*Please cite as:*
*Coumans JMJ, Oenema A, Bolman CAW, Lechner L*
*Use and Appreciation of a Web-Based, Computer-Tailored Diet and Physical Activity Intervention Based on the Self-determination Theory: Evaluation Study of Process and Predictors*
*JMIR Form Res 2021;5(12):e22390*
*URL: https://formative.jmir.org/2021/12/e22390*
*doi:10.2196/22390*
*PMID:34860670*

©Juul M J Coumans, Anke Oenema, Catherine A W Bolman, Lilian Lechner. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Development of a Web-Based School Support System Within the AVATAR Project for Psychosocial Well-being in Adolescents: Pilot Feasibility Study

Francesca Mastorci[1], PhD; Paolo Piaggi[2], PhD; Gabriele Trivellini[1]; Cristina Doveri[1]; Anselmo Casu[1]; Luca Bastiani[1], PhD; Irene Marinaro[1]; Cristina Vassalle[3], PhD; Alessandro Pingitore[1], MD, PhD

[1]Clinical Physiology Institute, Consiglio Nazionale delle Ricerche Area della Ricerca di Pisa (CNR), Pisa, Italy

[2]Department of Information Engineering, University of Pisa, Pisa, Italy

[3]Fondazione Toscana Gabriele Monasterio, Pisa, Italy

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

## Abstract

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

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

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

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

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



XSL•FO
RenderX

*(JMIR Form Res 2021;5(12):e24840)* doi:10.2196/24840

**KEYWORDS**

adolescent; well-being management; schools; web tool; health promotion

## Introduction

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

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

## Methods

### Overview

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

### General Aspects of the AVATAR Web-Based Tool

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

### Data Collection

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



XSL•FO
RenderX

levels were assessed using the Physical Activity Questionnaire for Older Children (PAQ-C). The questionnaire provides a general measure of physical activity for those aged 8 to 20 years. The PAQ-C is a self-administrated questionnaire consisting of 9 items rated on a 5-point scale. A higher score indicates more active children/adolescents [13].

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

## Health-Related Quality of Life Components

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

From our previous experimental evidence, the *lifestyle habits* component was hypothesized to cause changes in physical well-being, autonomy, financial resources, diet, and physical activity, while the *social context* component was hypothesized to cause changes in parent relations, peers, school environment, and bullying (analogous schemas were adopted for the social context and mental components).

Conceptually, the AVATAR methodology is based on a multidimensional construct, covering the physical, emotional, mental, and social components of well-being as perceived by adolescents [6,7,14,15]. The indicators belonging to the 4 components have been selected according to the analysis of pre-existing literature involving adolescents' health and well-being [16-18]. Table 1 presents the individual variables (observed variables) for each component.

**Table 1.** Variables (observed variables) for each component.

| Component | Well-being dimensions (observed variables) |
|---|---|
| Lifestyle habits | Physical well-being |
| | Autonomy |
| | Financial resources |
| | Diet |
| | Physical activity |
| Social context | Parent relations |
| | Peers |
| | School environment |
| | Bullying |
| Emotional status | Psychological well-being |
| | Mood |
| | Self-perception |
| | Emotion |

## Data Management

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

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

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



XSL•FO
RenderX

**Figure 1.** Example of data view. Results are expressed as being merged and integrated into a component (eg, lifestyle) or single variables of the entire class.

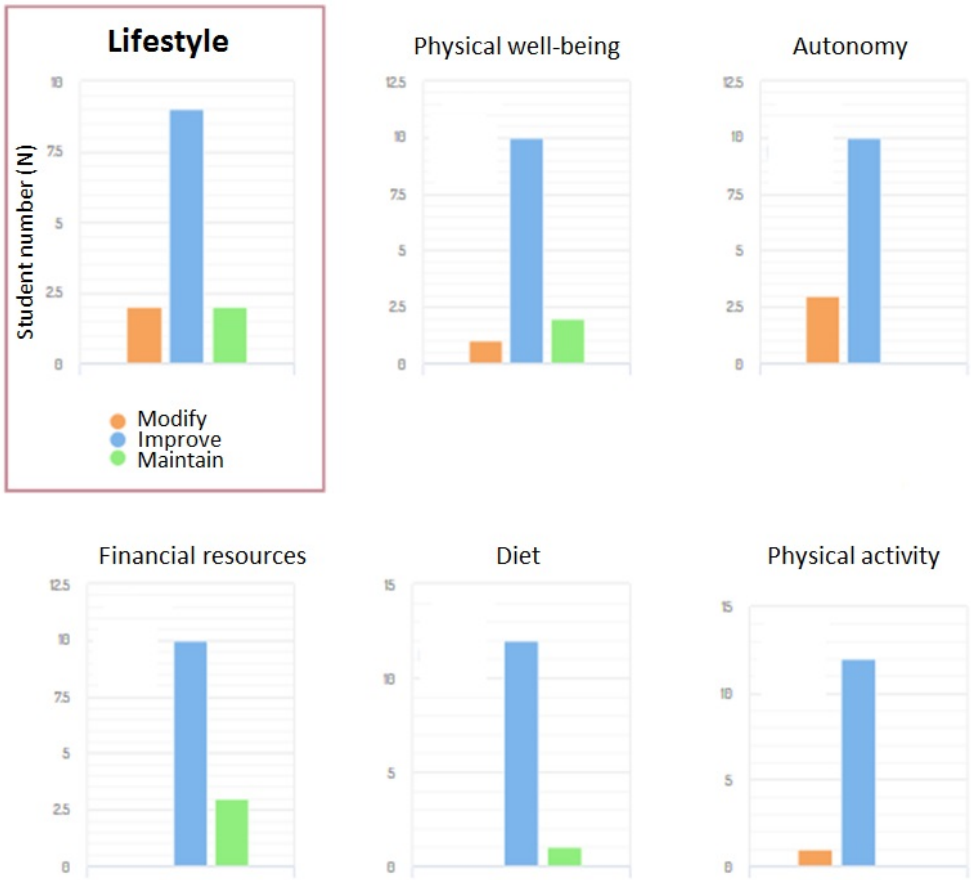
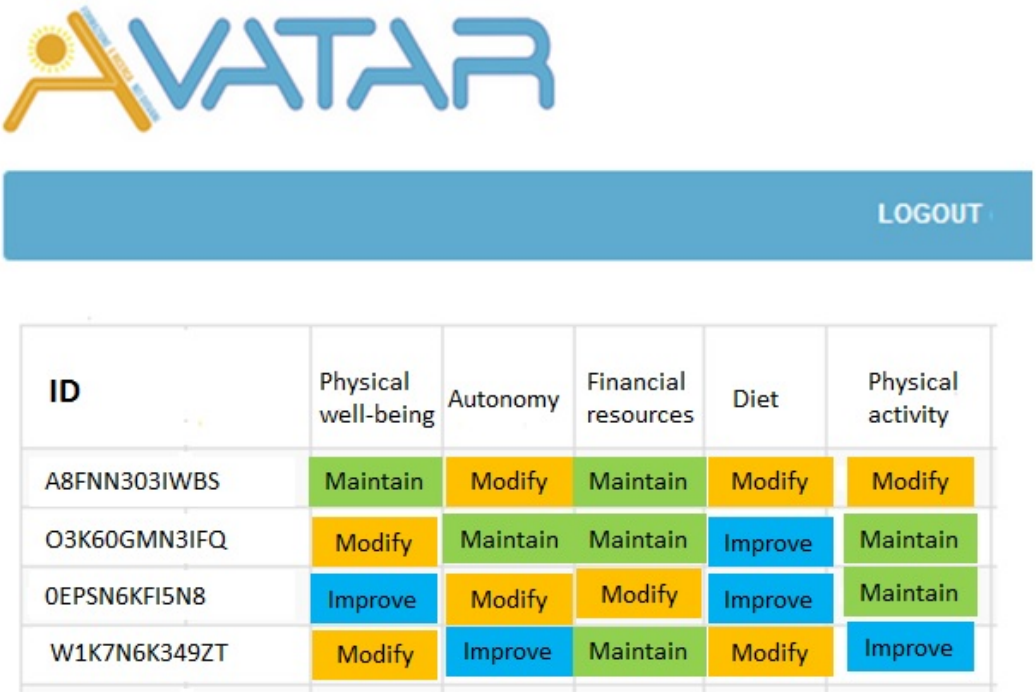



**Figure 2.** Details of the individual variables of a component (eg, lifestyle) for each student as shown on the platform classified as modify, improve, and maintain.




## Methods of the Pilot/Feasibility Study

In order to evaluate the feasibility of the AVATAR web-based school support system, a pilot study, as a part of the AVATAR project, was performed. Data collection was conducted between 2018 and 2019 from 1 of 10 junior high schools at the beginning (pre) and at the end (post) of the school year. In total, 331 students (172 female students, 52%; 159 male students, 48%;



XSL•FO
RenderX

mean age 12.5 years, SD 1 year) were included. Adolescent students were enrolled according to the following inclusion criteria: age 10 to 14 years, absence of neuropsychiatric or other diseases, informed consent signed, and completion of the questionnaires proposed. In every school class, all adolescents filled out the questionnaires, and those who were not eligible were excluded from the study retrospectively. The questionnaires, which have been previously described, were filled at the beginning and end of the school year.

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

### School-Based Program Built on the AVATAR Platform

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

### Statistical Analysis

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

## Results

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

Descriptive data on health-related quality of life and lifestyle habits (diet and physical activity; from preprogram and postprogram) are presented in Table 2. Data on the KIDSCREEN-52 dimensions are calculated as mean T-scores according to the KIDSCREEN Group [9,10]. During the school year, following the program built on the results of the first administration, students showed a higher perception in psychological well-being (*P*=.001), mood (*P*=.001), self-perception (*P*=.006), and autonomy (*P*=.001), which was understood as the opportunity to create his/her social and leisure time. There was an increase in the well-being perception owing to the school-based program, as well as the perception of financial resources compared with the initial situation (*P*=.001). In the social context, students reported higher values in the relationship with their family (*P*=.02) and peers (*P*=.001), and exhibited a lower perception of bullying (*P*=.001). For lifestyle habits, after the personalized program, students developed higher adherence to the Mediterranean diet (*P*=.007) and higher physical activity levels (*P*=.001) compared with the previous condition.



XSL•FO
RenderX

**Table 2.** Questionnaire findings in the pilot study sample at the beginning (pre) and end (post) of the school year.

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

[a]*P* values were calculated using the Student paired *t* test.

[b]Data on the KIDSCREEN-52 dimensions were calculated as mean T-scores according to the KIDSCREEN Group.

[c]KIDMED: Mediterranean Diet Quality Index for children and adolescents.

[d]PAQ-C: Physical Activity Questionnaire for Older Children.

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

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



XSL•FO
RenderX

**Table 3.** Changes in questionnaire findings (pre vs post) according to the modify, improve, and maintain percentile categories.

| Variable | Modify percentile category, % | | Improve percentile category, % | | Maintain percentile category, % | | Change (Δ), % | | | *P* value[a] |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pre | Post | Pre | Post | Pre | Post | Modify percentile | Improve percentile | Maintain percentile | |
| **KIDSCREEN-52** | | | | | | | | | | |
| Physical well-being | 18.4 | 18.1 | 67.4 | 50.0 | 14.2 | 23.9 | −0.3 | −9.4 | 9.7 | .001 |
| Psychological well-being | 27.5 | 7.6 | 66.8 | 78.5 | 5.7 | 13.9 | −19.9 | 11.8 | 8.2 | .001 |
| Mood/emotion | 8.8 | 5.4 | 78.5 | 76.1 | 12.7 | 18.4 | −3.3 | −2.4 | 5.7 | .03 |
| Self-perception | 12.1 | 10.3 | 87.9 | 89.7 | 0.0 | 0.0 | −1.8 | 1.8 | 0.0 | .42 |
| Autonomy | 10.6 | 4.5 | 77.3 | 72.5 | 12.1 | 23.0 | −6.0 | −4.8 | 10.9 | .001 |
| Parent relationship | 8.8 | 8.2 | 91.2 | 91.8 | 0.0 | 0.0 | −0.6 | 0.6 | 0.0 | .89 |
| Financial resources | 13.3 | 4.5 | 86.7 | 95.5 | 0.0 | 0.0 | −8.8 | 8.8 | 0.0 | .001 |
| Peers | 10.0 | 6.3 | 82.2 | 78.2 | 7.9 | 15.4 | −3.6 | −3.9 | 7.6 | .003 |
| School environment | 10.3 | 6.9 | 77.9 | 78.2 | 11.8 | 14.8 | −3.3 | 0.3 | 3.0 | .12 |
| Social acceptance | 13.3 | 7.9 | 86.7 | 92.1 | 0.0 | 0.0 | −5.4 | 5.4 | 0.0 | .01 |
| KIDMED[b] | 9.4 | 7.3 | 76.7 | 78.2 | 13.9 | 14.5 | −2.1 | 1.5 | 0.6 | .44 |
| PAQ-C[c] | 17.8 | 12.4 | 80.4 | 86.4 | 1.8 | 3.0 | −5.4 | 4.2 | 1.2 | .07 |

[a]*P* value is for the change (Δ).

[b]KIDMED: Mediterranean Diet Quality Index for children and adolescents.

[c]PAQ-C: Physical Activity Questionnaire for Older Children.

# Discussion

## Principal Findings

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

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

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

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



XSL•FO
RenderX

developed higher adherence to a Mediterranean diet and a better physical activity level.

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

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

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

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

## Strengths and Limitations

A key strength of the AVATAR platform is that it is the first platform, to our knowledge, designed to introduce a web-based health promotion tool for the school community [26]. Importantly, monitoring the results allows the identification of the strength and fragile characteristics of each adolescent in order to define personalized educational programs. In particular, with the data, the school can (1) carry out longitudinal monitoring of students, (2) orient didactic actions, and (3) promote the personalization of educational and training courses. In this context, the AVATAR platform intervenes at multiple points in the educational process in young people by combining management and empowerment of health and well-being, applying prevention strategies to reduce disease burden and health expenditure in adulthood, and enhancing learning. The AVATAR platform combines the rigor and objectivity of scientific and technological research with the needs that emerge from the school system regarding health, well-being, and educational success, integrating primordial prevention with the definition of a model of support for the autonomous design of schools. The development of a school-based program from the AVATAR web-based platform represents a powerful pattern for the promotion of health and psychosocial well-being, in which a network of different stakeholders dedicated to education may cooperate together to increase awareness, reduce risk behaviors, and potentiate educational success. Moreover, the AVATAR web-based tool offers a personalized well-being index that may allow the adoption of more individually focused strategies and interventions to improve well-being. Finally, as the AVATAR platform is delivered to students via web-based technology with interactive components through the involvement of teachers, student engagement and program fidelity can be increased, and this is supported by new national and European policies on the welfare of adolescents.

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

## Conclusion

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

## Acknowledgments

We thank the Italian schools that joined the Italian school network "RETE ULISSE, Scuole insieme per la ricerca scientifica e l'innovazione didattica." We particularly thank Istituto Comprensivo P Aldi (Grosseto) for pilot study monitoring.



XSL•FO
RenderX

## Conflicts of Interest

None declared.

## References


1. Jones SM, Bouffard SM. Social and Emotional Learning in Schools: From Programs to Strategies and commentaries. Social Policy Report 2012 Dec 01;26(4):1-33. [doi: 10.1002/j.2379-3988.2012.tb00073.x]
2. Weare K, Nind M. Mental health promotion and problem prevention in schools: what does the evidence say? Health Promot Int 2011 Dec 12;26 Suppl 1(suppl 1):i29-i69. [doi: 10.1093/heapro/dar075] [Medline: 22079935]
3. Buttitta M, Iliescu C, Rousseau A, Guerrien A. Quality of life in overweight and obese children and adolescents: a literature review. Qual Life Res 2014 May 19;23(4):1117-1139. [doi: 10.1007/s11136-013-0568-5] [Medline: 24249217]
4. Ross DA, Hinton R, Melles-Brewer M, Engel D, Zeck W, Fagan L, et al. Adolescent Well-Being: A Definition and Conceptual Framework. J Adolesc Health 2020 Oct;67(4):472-476 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.06.042] [Medline: 32800426]
5. Ravens-Sieberer U, Herdman M, Devine J, Otto C, Bullinger M, Rose M, et al. The European KIDSCREEN approach to measure quality of life and well-being in children: development, current application, and future advances. Qual Life Res 2014 Apr 18;23(3):791-803 [FREE Full text] [doi: 10.1007/s11136-013-0428-3] [Medline: 23686556]
6. Mastorci F, Bastiani L, Doveri C, Trivellini G, Casu A, Vassalle C, et al. Adolescent Health: A Framework for Developing an Innovative Personalized Well-Being Index. Front Pediatr 2020 May 7;8:181-189 [FREE Full text] [doi: 10.3389/fped.2020.00181] [Medline: 32457853]
7. Trivellini G, Doveri C, Mastorci F, Bastiani L, Cappa C, Vassalle C, et al. Innovative web-based tool for promoting well-being among healthy adolescents: An implementation protocol. J Transl Sci 2018;5(3):A. [doi: 10.15761/JTS.1000261]
8. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. Bull World Health Organ 2007 Sep 01;85(9):660-667 [FREE Full text] [doi: 10.2471/blt.07.043497] [Medline: 18026621]
9. Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Duer W, European Kidscreen Group. KIDSCREEN-52 quality-of-life measure for children and adolescents. Expert Rev Pharmacoecon Outcomes Res 2005 Jun 09;5(3):353-364. [doi: 10.1586/14737167.5.3.353] [Medline: 19807604]
10. The KIDSCREEN Group Europe. The KIDSCREEN Questionnaires - Quality of life questionnaires for children and adolescents. Lengerich, Germany: Pabst Science Publishers; 2006.
11. Berra S, Ravens-Sieberer U, Erhart M, Tebé C, Bisegger C, Duer W, European KIDSCREEN group [kidscreen]. Methods and representativeness of a European survey in children and adolescents: the KIDSCREEN study. BMC Public Health 2007 Jul 26;7(1):182-194 [FREE Full text] [doi: 10.1186/1471-2458-7-182] [Medline: 17655756]
12. Serra-Majem L, Ribas L, Ngo J, Ortega RM, García A, Pérez-Rodrigo C, et al. Food, youth and the Mediterranean diet in Spain. Development of KIDMED, Mediterranean Diet Quality Index in children and adolescents. Public Health Nutr 2004 Oct 02;7(7):931-935. [doi: 10.1079/phn2004556] [Medline: 15482620]
13. Saint-Maurice PF, Welk GJ, Beyler NK, Bartee RT, Heelan KA. Calibration of self-report tools for physical activity research: the Physical Activity Questionnaire (PAQ). BMC Public Health 2014 May 16;14(1):461-470 [FREE Full text] [doi: 10.1186/1471-2458-14-461] [Medline: 24886625]
14. Mastorci F, Piaggi P, Bastiani L, Trivellini G, Doveri C, Casu A, et al. The impact of menarche on health-related quality of life in a sample of Italian adolescents: evidence from school-based AVATAR project. Eur J Pediatr 2020 Jun 04;179(6):973-978. [doi: 10.1007/s00431-020-03594-8] [Medline: 32020330]
15. Mastorci F, Bastiani L, Trivellini G, Doveri C, Vassalle C, Pingitore A. A new integrated approach for adolescent health and well-being: the AVATAR project. Health Qual Life Outcomes 2020 Mar 18;18(1):77-86 [FREE Full text] [doi: 10.1186/s12955-020-01291-6] [Medline: 32188459]
16. Patton GC, Coffey C, Cappa C, Currie D, Riley L, Gore F, et al. Health of the world's adolescents: a synthesis of internationally comparable data. The Lancet 2012 Apr;379(9826):1665-1675. [doi: 10.1016/s0140-6736(12)60203-7]
17. Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. The Lancet 2016 Jun;387(10036):2423-2478. [doi: 10.1016/s0140-6736(16)00579-1]
18. Lamblin M, Murawski C, Whittle S, Fornito A. Social connectedness, mental health and the adolescent brain. Neurosci Biobehav Rev 2017 Sep;80:57-68. [doi: 10.1016/j.neubiorev.2017.05.010] [Medline: 28506925]
19. Rajmil L, Alonso J, Berra S, Ravens-Sieberer U, Gosch A, Simeoni M, KIDSCREEN group. Use of a children questionnaire of health-related quality of life (KIDSCREEN) as a measure of needs for health care services. J Adolesc Health 2006 May;38(5):511-518. [doi: 10.1016/j.jadohealth.2005.05.022] [Medline: 16635761]
20. Berman AH, Liu B, Ullman S, Jadbäck I, Engström K. Children's Quality of Life Based on the KIDSCREEN-27: Child Self-Report, Parent Ratings and Child-Parent Agreement in a Swedish Random Population Sample. PLoS One 2016 Mar 9;11(3):e0150545 [FREE Full text] [doi: 10.1371/journal.pone.0150545] [Medline: 26959992]
21. Elias M, Weissberg R. Primary prevention: educational approaches to enhance social and emotional learning. J Sch Health 2000 May;70(5):186-190. [doi: 10.1111/j.1746-1561.2000.tb06470.x] [Medline: 10900595]




XSL•FO
RenderX

22. Perrin JM, Duncan G, Diaz A, Kelleher K. Principles And Policies To Strengthen Child And Adolescent Health And Well-Being. Health Aff (Millwood) 2020 Oct 01;39(10):1677-1683. [doi: 10.1377/hlthaff.2020.00709] [Medline: 33017241]
23. Weissberg RP, Elias MJ. Enhancing young people's social competence and health behavior: An important challenge for educators, scientists, policymakers, and funders. Applied and Preventive Psychology 1993;2(4):179-190. [doi: 10.1016/s0962-1849(05)80088-5]
24. Stice E, Becker CB, Yokum S. Eating disorder prevention: current evidence-base and future directions. Int J Eat Disord 2013 Jul 09;46(5):478-485 [FREE Full text] [doi: 10.1002/eat.22105] [Medline: 23658095]
25. Xu T, Tomokawa S, Gregorio ER, Mannava P, Nagai M, Sobel H. School-based interventions to promote adolescent health: A systematic review in low- and middle-income countries of WHO Western Pacific Region. PLoS One 2020 Mar 5;15(3):e0230046 [FREE Full text] [doi: 10.1371/journal.pone.0230046] [Medline: 32134985]
26. Langford R, Bonell C, Jones H, Pouliou T, Murphy S, Waters E, et al. The World Health Organization's Health Promoting Schools framework: a Cochrane systematic review and meta-analysis. BMC Public Health 2015 Mar 12;15(1):130-145 [FREE Full text] [doi: 10.1186/s12889-015-1360-y] [Medline: 25886385]

## Abbreviations

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

*Edited by G Eysenbach; submitted 08.10.20; peer-reviewed by D Vollmer Dahlke, E Sakellari; comments to author 26.02.21; revised version received 31.05.21; accepted 06.07.21; published 02.12.21.*

*Please cite as:*
*Mastorci F, Piaggi P, Trivellini G, Doveri C, Casu A, Bastiani L, Marinaro I, Vassalle C, Pingitore A*
*Development of a Web-Based School Support System Within the AVATAR Project for Psychosocial Well-being in Adolescents: Pilot Feasibility Study*
*JMIR Form Res 2021;5(12):e24840*
*URL: https://formative.jmir.org/2021/12/e24840*
*doi:10.2196/24840*
*PMID:34860668*

©Francesca Mastorci, Paolo Piaggi, Gabriele Trivellini, Cristina Doveri, Anselmo Casu, Luca Bastiani, Irene Marinaro, Cristina Vassalle, Alessandro Pingitore. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Development of the Shift Smartphone App to Support the Emotional Well-Being of Junior Physicians: Design of a Prototype and Results of Usability and Acceptability Testing

Isabelle Counson[1,2], MPH, GradDipPsychAdv; Alexandra Bartholomew[1], BCom (BSc), MPH; Joanna Crawford[1,2], PhD; Katherine Petrie[1,2], BSc (Hons); Geetanjali Basarkod[3], PhD, BSc (Hons); Victoria Moynihan[1], BVA; Josie Pires[1], BSc; Rachel Cohen[1], PhD, BSc (Hons), MPsych; Nicholas Glozier[4], MBBS, FRANZCP, PhD; Samuel Harvey[1,2], MBBS, MRCGP, MRCPsych, FRANZCP, PhD; Samineh Sanatkar[1,2], PhD

[1]Black Dog Institute, Randwick, Australia

[2]School of Psychiatry, Faculty of Medicine, UNSW Sydney, Kensington, Australia

[3]Institute for Positive Psychology and Education, Australian Catholic University, North Sydney, Australia

[4]Brain and Mind Centre, Faculty of Medicine and Health, University of Sydney, Australia

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

## Abstract

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

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

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

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

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

***(JMIR Form Res 2021;5(12):e26370)*** doi:10.2196/26370



XSL•FO
RenderX

**KEYWORDS**

digital mental health; mHealth apps; help-seeking; junior physicians; co-design; user-centered design; mobile phone

## Introduction

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

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

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

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

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

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

## Methods

### Overview

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



XSL•FO
RenderX

available to our team was to utilize the *30-day challenge* format of the evidence-based *HeadGear* app and to modify the clinical content to meet the needs of junior physicians. To this end, the interviews with junior physicians in phase 1 included a question about their attitudes toward a *30-day challenge* in an app to support the well-being of junior physicians.

### Phase 1: Needs Assessment Through Qualitative Interviews

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

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

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

### Phase 2: Prototype Development

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

### Phase 3: Pilot-Testing

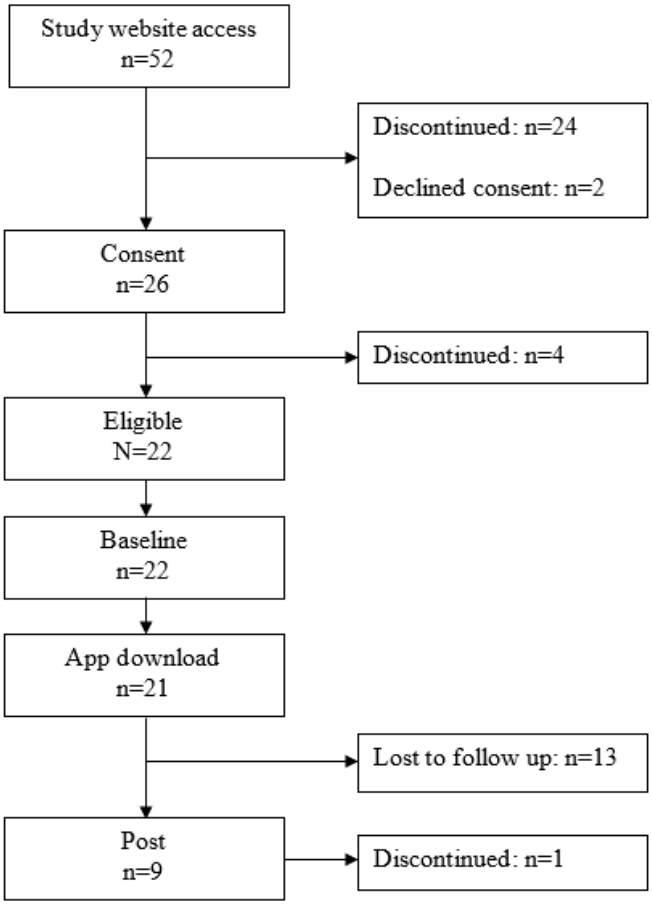
Pilot-testing was conducted in October and November 2019 to examine the usability and acceptability of the *prototype version of the Shift app*. Trends in depression and anxiety symptom severity, as well as changes in help-seeking intentions before and after using the *Shift* app prototype over a 4-week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

Two New South Wales hospital sites, one regional and one metropolitan, issued recruitment calls via email messages on behalf of the research team to junior physicians at the intern, resident, or registrar levels. The eligibility criteria were current employment as junior medical officers in New South Wales and ownership of an internet-enabled smartphone with an Apple or Android operating system. A total of 52 candidates accessed the study website, of which 50% (26/52) consented to participate. A final sample of 22 participants (13/22, 59% women; mean age 29, SD 4.1 years) entered the pilot study and completed a baseline assessment. A diagram of the participant flow is shown in Figure 1.



XSL•FO
RenderX

**Figure 1.** Flow chart describing how participants progressed through the pilot study phases.




A baseline questionnaire assessed basic demographic information (age, gender, level of training, and regional or metropolitan placement) and symptoms of depression (Patient Health Questionnaire 9-item) and anxiety (General Anxiety Disorder scale 7-item) over the past 2 weeks. Participants indicated their depression and anxiety symptoms on a 4-point Likert-type response scale ranging from 0 (*not at all*) to 3 (*nearly every day*). In addition, participants indicated their previous and recent help-seeking intentions for mental health problems (eg, "If you were to develop a mental health issue, how likely would you be to seek help from a GP or mental health professional [eg, psychologist/psychiatrist]?") on an 11-point Likert-type response scale ranging from 0 (*not at all likely*) to 10 (*very likely*). After completion of the baseline questionnaire, participants were given instructions on how to download the *Shift* app onto their mobile phones. The app automatically recorded usage metrics, such as the number of log-ins and challenges or sessions completed.

After 1 month, participants were invited via email to complete a poststudy questionnaire reassessing their current depression and anxiety symptom levels and help-seeking intentions. At the poststudy assessment point, participants were further asked to respond to a battery of purpose-built questions relating to the usability and acceptability of the *Shift* app prototype (eg, "Was the app interesting/engaging?") on a 5-point Likert-type scale ranging from 1 (*not at all*) to 5 (*completely*) and to the overall satisfaction rating of the app on a 5-point Likert-type scale ranging from 1 (*low*) to 5 (*high*). A modified version of the System Usability Scale was also administered to obtain an objective indication of the overall ease of use of the app.

### Phase 4: Shift App Redevelopment

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

## Results

### Phase 1: Needs Assessment Through Qualitative Interviews

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

All participants reported that they owned a smartphone and used apps multiple times each day. Half of the sample had already tried at least one mindfulness app, and 2 had previously used a web-based mental health app. Although the use of work-related apps was common, the use of such apps was centered on communication and medical information–seeking and reference



XSL•FO
RenderX

material. A total of 83% (10/12) participants reported to be *very or somewhat interested* in the idea of an app to support the mental health of junior physicians and provided positive endorsement for the idea in principle and for its potential to benefit the population. Most of the participants endorsed use of the app for both prevention of and early intervention in mental illness and suggested the inclusion of directions to support and treatment services for those with more severe mental illness. Although no participants were directly opposed to the idea of an mHealth app to support junior physicians' mental health, 2 participants questioned whether fellow junior physicians would use an mHealth app or to what extent they would benefit from it, particularly if physicians were not currently experiencing symptoms of mental illness ("no one takes their medicine when they are feeling good").

### App Naming, Content, and Feature Suggestions

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

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

**Table 1.** Qualitative interview participants' overview of main content recommendations for *Shift* and whether recommendations were adopted in successive versions of the app (N=12).

| Recommended content components | Version 1 | Version 2 |
|---|---|---|
| Cognitive behavioral therapy | ✓ | ✓[a] |
| Sleep hygiene | ✓ | ✓ |
| Mindfulness and stress management | ✓ | ✓ |
| Goal-setting | ✓ | ✓ |
| Pleasant activity scheduling | ✓ | ✓ |
| Practical lifestyle strategies | ✓ | ✓[a] |
| Problematic work situations | ✓ | ✓[a] |
| Hand over tips for changing terms or hospitals | | |
| Stories from junior physicians | | ✓[a] |

[a]Indicates an addition or improvement compared with the previous version.

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



XSL•FO
RenderX

**Table 2.** Qualitative interview participants' desired features for *Shift* and whether recommendations were adopted (N=12).

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

[a]Indicates improvement compared with the previous version.

### *Facilitators of and Barriers to App Use*

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

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

## Phase 2: Prototype Development

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

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

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

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

The clinical content and design drafts of the app prototype were modified based on user experience feedback. User experience experts and psychologists worked with 2 junior physicians to refine the user pathways, functionality, color palette, and design and to modify the clinical content (eg, examples of scenarios used in cognitive therapy). Tables 3 and 4 outline the contents of the resulting *Shift* app prototype, which were organized into challenges and sessions.



XSL•FO
RenderX

**Table 3.** Organization of challenge topics in the *Shift* app prototype version.

| Therapeutic type and challenge day or days | Topic name | Type |
|---|---|---|
| **Mindfulness** | | |
| 2 | Introduction to mindfulness | Video |
| 3 | Seeing the horizon | Audios |
| 9 | Grounding anchor | Audios |
| 16 | Calming breath | Audios |
| 20 | Loving-kindness | Audios |
| 24 | Cargo thoughts | Audios |
| 26 | Breathing wind | Audios |
| 28 | Lapping ocean | Audios |
| **Value-based** | | |
| 4 | Introduction to values and values as a physician | Video |
| 5 | Strive for five | Text |
| 6, 12, 17, 21, 27 | Scheduling meaningful actions | Text |
| **Cognitive behavioral** | | |
| 7 | Introduction to unhelpful thoughts | Video |
| 8 | Unhelpful thoughts | Text |
| 11 | Cognitive biases | Text |
| 14 | Introduction to thought challenging | Video |
| 15 | Thought challenging | Text |
| 22 | Worry decision tree | Text |
| 23 | Cognitive therapy review | Video |
| **Positive psychology** | | |
| 10 | Gratitude | Text |
| 13 | Getting active | Text |
| 18 | Social support | Text |
| 19 | Help a friend | Text |
| 25 | 100 enjoyable activities | Text |
| 29 | Planning for the future | Text |
| **General** | | |
| 1 | Checkup | Text |
| 30 | Putting it all together | Video |



XSL•FO
RenderX

**Table 4.** Organization of session topics in the *Shift* app prototype version.

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

[a]JMO: junior medical officer.

## Phase 3: Pilot-Testing

### *App Acceptability and Usability*

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

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

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

[a]Response scales ranged from 1 to 5.



XSL•FO
RenderX

**Table 6.** Pilot trial participants' usefulness ratings of the *Shift* app prototype 30-day challenge contents (N=8)[a].

| Challenge type and topic name[b] | Values, median (range) | Values, minimum-maximum |
|---|---|---|
| Mindfulness | 4 (5) | 0-5 |
| Value-based | 0 (5) | 0-5 |
| Cognitive behavioral | 3 (5) | 0-5 |
| **Positive psychology** | | |
| Gratitude | 3 (5) | 0-5 |
| Getting active | 1.5 (5) | 0-5 |
| Social support | 2.5 (5) | 0-5 |
| Help a friend | 0 (4) | 0-4 |
| Enjoyable activities | 3 (5) | 0-5 |
| Planning for the future | 1.5 (5) | 0-5 |
| General | 0 (4) | 0-4 |

[a]Response scales ranged from 0 to 5, where 0 indicates unattempted components, 1 indicates low perceived usefulness, and 5 indicates high perceived usefulness.

[b]Only positive psychology challenge topics were assessed individually because of the distinctiveness of each topic in this category.

**Table 7.** Pilot trial participants' usefulness ratings of the *Shift* app prototype session contents (N=8)[a].

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

[a]Response scales ranged from 0 to 5, where 0 indicates unattempted components, 1 indicates low perceived usefulness, and 5 indicates high perceived usefulness.

[b]JMO: junior medical officer.



XSL•FO
RenderX

### App Use

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

### Symptoms Change

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

## Phase 4: Shift App Redevelopment

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

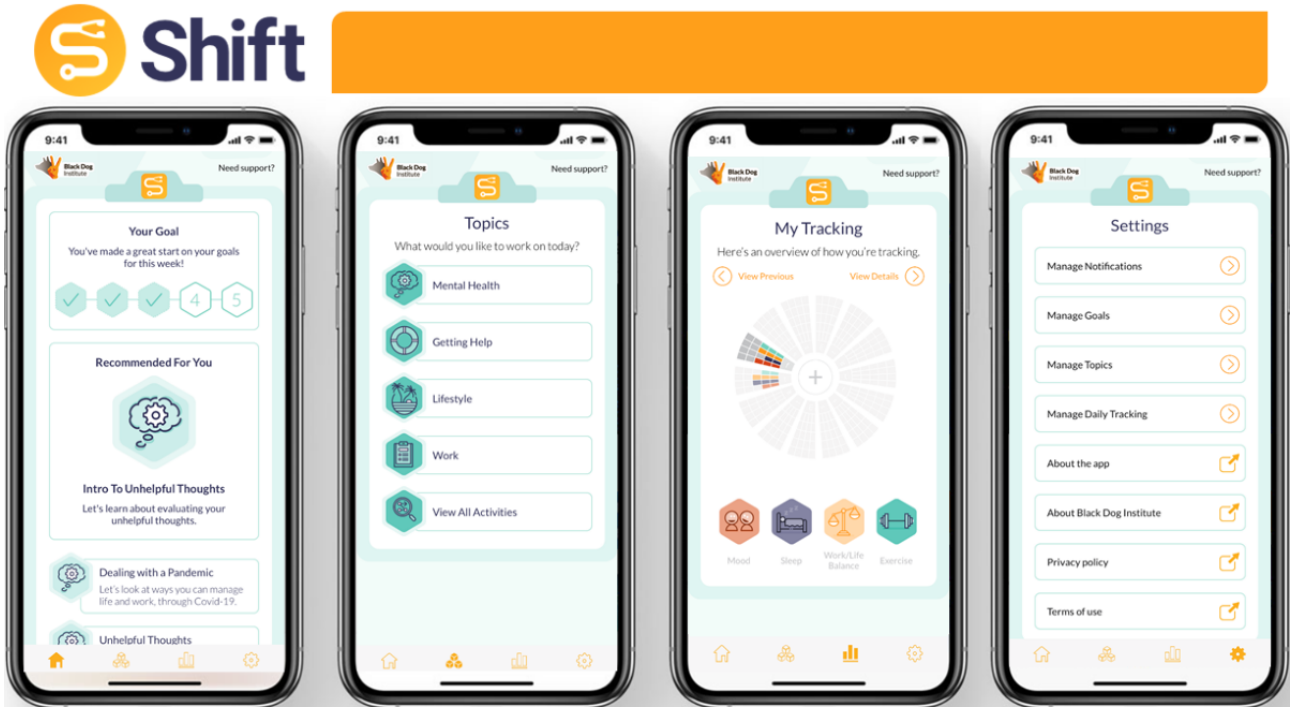
A streamlined login process with the inclusion of a biometric security system (ie, fingerprint or face ID authentication) was incorporated to facilitate use after app download. Personalization enhancements were achieved through significant changes in the presentation of the app. Importantly, the 30-day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

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



XSL•FO
RenderX

**Figure 2.** Visual examples of the most recent *Shift* version Home, Topics, Tracking, and Settings screens.




## Discussion

### Principal Findings

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

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

### Strengths and Limitations

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



XSL•FO
RenderX

put forward by junior physicians into an mHealth app intervention environment.

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

## Conclusions

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

## Acknowledgments

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

## Conflicts of Interest

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

## References

1. Markwell AL, Wainer Z. The health and wellbeing of junior doctors: insights from a national survey. Med J Aust 2009 Oct 19;191(8):441-444. [doi: 10.5694/j.1326-5377.2009.tb02880.x] [Medline: 19835538]
2. Wu F, Ireland M, Hafekost K, Lawrence D. National mental health survey of doctors and medical students. Beyond Blue. URL: https://www.beyondblue.org.au/docs/default-source/research-project-files/bl1132-report---nmhdmss-full-report_web [accessed 2021-11-08]
3. Forbes MP, Iyengar S, Kay M. Barriers to the psychological well-being of Australian junior doctors: a qualitative analysis. BMJ Open 2019 Jun 12;9(6):e027558 [FREE Full text] [doi: 10.1136/bmjopen-2018-027558] [Medline: 31196900]
4. Medical training survey. Medical Board Aphra. URL: https://www.medicalboard.gov.au/Registration/Medical-training-survey.aspx [accessed 2021-11-08]
5. Tyssen R, Vaglum P. Mental health problems among young doctors: an updated review of prospective studies. Harv Rev Psychiatry 2002;10(3):154-165. [doi: 10.1080/10673220216218] [Medline: 12023930]
6. Brooks SK, Del Busso L, Chalder T, Harvey SB, Hatch SL, Hotopf M, et al. 'You feel you've been bad, not ill': sick doctors' experiences of interactions with the General Medical Council. BMJ Open 2014 Jul 17;4(7):e005537 [FREE Full text] [doi: 10.1136/bmjopen-2014-005537] [Medline: 25034631]
7. Brooks SK, Gerada C, Chalder T. Review of literature on the mental health of doctors: are specialist services needed? J Ment Health 2011 Apr;20(2):146-156. [doi: 10.3109/09638237.2010.541300] [Medline: 21275504]
8. Wallace JE. Mental health and stigma in the medical profession. Health (London) 2012 Jan;16(1):3-18. [doi: 10.1177/1363459310371080] [Medline: 21177717]
9. Muhamad Ramzi NS, Deady M, Petrie K, Crawford J, Harvey S. Help-seeking for depression among Australian doctors. Intern Med J 2020 Aug 24:- (forthcoming). [doi: 10.1111/imj.15035] [Medline: 32833296]



XSL•FO
RenderX

10. Brooks SK, Chalder T, Gerada C. Doctors vulnerable to psychological distress and addictions: treatment from the Practitioner Health Programme. J Ment Health 2011 Apr;20(2):157-164. [doi: 10.3109/09638237.2011.556168] [Medline: 21542716]
11. Fahrenkopf AM, Sectish TC, Barger LK, Sharek PJ, Lewin D, Chiang VW, et al. Rates of medication errors among depressed and burnt out residents: prospective cohort study. BMJ 2008 Feb 07;336(7642):488-491. [doi: 10.1136/bmj.39469.763218.be]
12. West CP, Tan AD, Habermann TM, Sloan JA, Shanafelt TD. Association of resident fatigue and distress with perceived medical errors. JAMA 2009 Sep 23;302(12):1294-1300. [doi: 10.1001/jama.2009.1389] [Medline: 19773564]
13. Williams ES, Manwell LB, Konrad TR, Linzer M. The relationship of organizational culture, stress, satisfaction, and burnout with physician-reported error and suboptimal patient care: results from the MEMO study. Health Care Manage Rev 2007;32(3):203-212. [doi: 10.1097/01.HMR.0000281626.28363.59] [Medline: 17666991]
14. Petrie K, Crawford J, Baker ST, Dean K, Robinson J, Veness BG, et al. Interventions to reduce symptoms of common mental disorders and suicidal ideation in physicians: a systematic review and meta-analysis. Lancet Psychiatry 2019 Mar;6(3):225-234. [doi: 10.1016/s2215-0366(18)30509-1]
15. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med 2018 Jun;283(6):516-529 [FREE Full text] [doi: 10.1111/joim.12752] [Medline: 29505159]
16. Guille C, Zhao Z, Krystal J, Nichols B, Brady K, Sen S. Web-based cognitive behavioral therapy intervention for the prevention of suicidal ideation in medical interns: a randomized clinical trial. JAMA Psychiatry 2015 Dec;72(12):1192-1198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1880] [Medline: 26535958]
17. Frögéli E, Djordjevic A, Rudman A, Livheim F, Gustavsson P. A randomized controlled pilot trial of acceptance and commitment training (ACT) for preventing stress-related ill health among future nurses. Anxiety Stress Coping 2016;29(2):202-218. [doi: 10.1080/10615806.2015.1025765] [Medline: 25759942]
18. Kushnir T, Malkinson R, Ribak J. Rational thinking and stress management in health workers: a psychoeducational program. Int J Stress Manag 1998;5(3):169-178. [doi: 10.1023/A:1022941031900]
19. Reeve A, Tickle A, Moghaddam N. Are acceptance and commitment therapy-based interventions effective for reducing burnout in direct-care staff? A systematic review and meta-analysis. Ment Health Rev J 2018 Jul 04;23(3):131-155. [doi: 10.1108/mhrj-11-2017-0052]
20. de Korte E, Wiezer N, Bakhuys Roozeboom M, Vink P, Kraaij W. Behavior change techniques in mHealth apps for the mental and physical health of employees: systematic assessment. JMIR Mhealth Uhealth 2018 Oct 03;6(10):e167 [FREE Full text] [doi: 10.2196/mhealth.6363] [Medline: 30282621]
21. Deady M, Johnston D, Milne D, Glozier N, Peters D, Calvo R, et al. Preliminary effectiveness of a smartphone app to reduce depressive symptoms in the workplace: feasibility and acceptability study. JMIR Mhealth Uhealth 2018 Dec 04;6(12):e11661 [FREE Full text] [doi: 10.2196/11661] [Medline: 30514694]
22. Ganesan AN, Louise J, Horsfall M, Bilsborough SA, Hendriks J, McGavigan AD, et al. International mobile-health intervention on physical activity, sitting, and weight: the stepathlon cardiovascular health study. J Am Coll Cardiol 2016 May 31;67(21):2453-2463 [FREE Full text] [doi: 10.1016/j.jacc.2016.03.472] [Medline: 27050185]
23. Johnson J, Devdutt J, Mehrotra S, Bhola P, Sudhir P, Sharma A. Barriers to professional help-seeking for distress and potential utility of a mental health app components: stakeholder perspectives. Cureus 2020 Feb 28;12(2):e7128 [FREE Full text] [doi: 10.7759/cureus.7128] [Medline: 32257673]
24. Hill C, Martin JL, Thomson S, Scott-Ram N, Penfold H, Creswell C. Navigating the challenges of digital health innovation: considerations and solutions in developing online and smartphone-application-based interventions for mental health disorders. Br J Psychiatry 2017 Aug;211(2):65-69. [doi: 10.1192/bjp.bp.115.180372] [Medline: 28522435]
25. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020 Jan 06;22(1):e15337 [FREE Full text] [doi: 10.2196/15337] [Medline: 31904579]
26. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
27. de Korte EM, Wiezer N, Janssen JH, Vink P, Kraaij W. Evaluating an mHealth app for health and well-being at work: mixed-method qualitative study. JMIR Mhealth Uhealth 2018 Mar 28;6(3):e72 [FREE Full text] [doi: 10.2196/mhealth.6335] [Medline: 29592846]
28. Melzner J, Heinze J, Fritsch T. Mobile health applications in workplace health promotion: an integrated conceptual adoption framework. Proc Technol 2014;16:1374-1382. [doi: 10.1016/j.protcy.2014.10.155]
29. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
30. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
31. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol (New York) 2016 Jun;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]




XSL•FO
RenderX

32. Deady M, Peters D, Lang H, Calvo R, Glozier N, Christensen H, et al. Designing smartphone mental health applications for emergency service workers. Occup Med (Lond) 2017 Aug 01;67(6):425-428. [doi: 10.1093/occmed/kqx056] [Medline: 28535246]
33. Peters D, Deady M, Glozier N, Harvey S, Calvo RA. Worker preferences for a mental health app within male-dominated industries: participatory study. JMIR Ment Health 2018 Apr 25;5(2):e30 [FREE Full text] [doi: 10.2196/mental.8999] [Medline: 29695371]
34. Petrie K, Crawford J, Deady M, Lupton D, Boydell K, Harvey S. 'The Hardest Job I've Ever Done?: a qualitative exploration of the factors affecting junior doctors' mental health and wellbeing during medical training in Australia. BMC Health; under review; Services Research 2021:- (forthcoming).
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Massey OT. A proposed model for the analysis and interpretation of focus groups in evaluation research. Eval Program Plann 2011 Feb;34(1):21-28. [doi: 10.1016/j.evalprogplan.2010.06.003] [Medline: 20655593]
37. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Human Comput Inter 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
38. Kortum P, Sorber M. Measuring the usability of mobile applications for phones and tablets. Int J Human Comput Inter 2015 Jul 31;31(8):518-529. [doi: 10.1080/10447318.2015.1064658]
39. Honey P, Mumford A. Manual of Learning Styles. Berkshire: Peter Honey; 1982.
40. Knowles M. Andragogy: An emerging technology for adult learning. In: The Modern Practice of Adult Education: From Pedagogy to Andragogy. London, UK: Cambridge Book Company; 1970.
41. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Comput Biol Med 2002 May;32(3):141-149. [doi: 10.1016/s0010-4825(02)00011-2] [Medline: 11922931]

## Abbreviations

**mHealth:** mobile health

*Edited by G Eysenbach; submitted 09.12.20; peer-reviewed by S Six, L Balcombe, A Knapp; comments to author 25.02.21; revised version received 09.04.21; accepted 21.10.21; published 02.12.21.*

*Please cite as:*
*Counson I, Bartholomew A, Crawford J, Petrie K, Basarkod G, Moynihan V, Pires J, Cohen R, Glozier N, Harvey S, Sanatkar S Development of the Shift Smartphone App to Support the Emotional Well-Being of Junior Physicians: Design of a Prototype and Results of Usability and Acceptability Testing*
*JMIR Form Res 2021;5(12):e26370*
*URL: https://formative.jmir.org/2021/12/e26370*
*doi:10.2196/26370*
*PMID:34860662*

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



XSL•FO
RenderX

Original Paper

# The Use of Food Images and Crowdsourcing to Capture Real-time Eating Behaviors: Acceptability and Usability Study

Katharine Harrington[1], MPH; Shannon N Zenk[2,3], PhD; Linda Van Horn[1], PhD; Lauren Giurini[4]; Nithya Mahakala[1]; Kiarri N Kershaw[1], PhD

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

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

## Abstract

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

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

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

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

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

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

**KEYWORDS**

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



XSL•FO
RenderX

## Introduction

### Background

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

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

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

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

### Objective

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

## Methods

### Study Population

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

### Study Procedure

The CHEERS EMA pilot study comprised a 1-day initial visit, a 7-day EMA study period, and a 1-day final visit. Participants were incentivized to participate with cash rewards: US $20 for the first visit, US $9 per day of valid data collection (up to US $63), and US $30 for the final visit. During the initial visit,



XSL•FO
RenderX

participants were asked to complete several questionnaires adopted from other studies or created specifically for CHEERS. Participants also completed the validated Block Fat/Sugar/Fruit/Vegetable Screener, which provides estimates of saturated fat, trans fat, total sugars, *added sugars* (in sweetened cereals, soft drinks, and sweets), fruit and fruit juice, vegetable intake, glycemic load, and glycemic index [29]. In this study, the Block Fat/Sugar/Fruit/Vegetable Screener was used to estimate the usual weekly intake of fruits, vegetables, sweet snacks, and salty snacks.

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

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

## Usual Food Intake

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



XSL•FO
RenderX

**Textbox 1.** Questions taken from the Block Fat/Sugar/Fruit/Vegetable Screener at previsit to determine the self-reported intake of fruit, vegetables, salty snacks, and sweet snacks.

**Fruit**

- Any kind of fruit, fresh or canned (not counting juice)

**Vegetables**

- Green salad and vegetables you put in green salad
- Potatoes, not fried, such as baked or mashed
- Vegetable soup or stew with vegetables
- All other vegetables you eat as a side dish or in any kind of dish, not counting salad or potatoes

**Salty snacks**

- French fries, home fries, and hash browns
- Snack chips such as potato chips, tortilla, corn chips, Fritos, Doritos, or popcorn (not pretzels)
- Crackers such Ritz, soda crackers, Cheez-Its, or any other snack cracker

**Sweet snacks**

- Ice cream and ice cream bars
- Donuts
- Cake, cookies, or snack cakes such as cupcakes, Twinkies, or any other pastry
- Pie, including fast food pies or snack pies
- Chocolate candy such as chocolate bars, M&Ms, Mars Bars, and Reeses
- Any other candy (not chocolate) such as hard candy, Lifesavers, Skittles, and Starburst

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

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

## Crowdsourced Labeling of Food Images

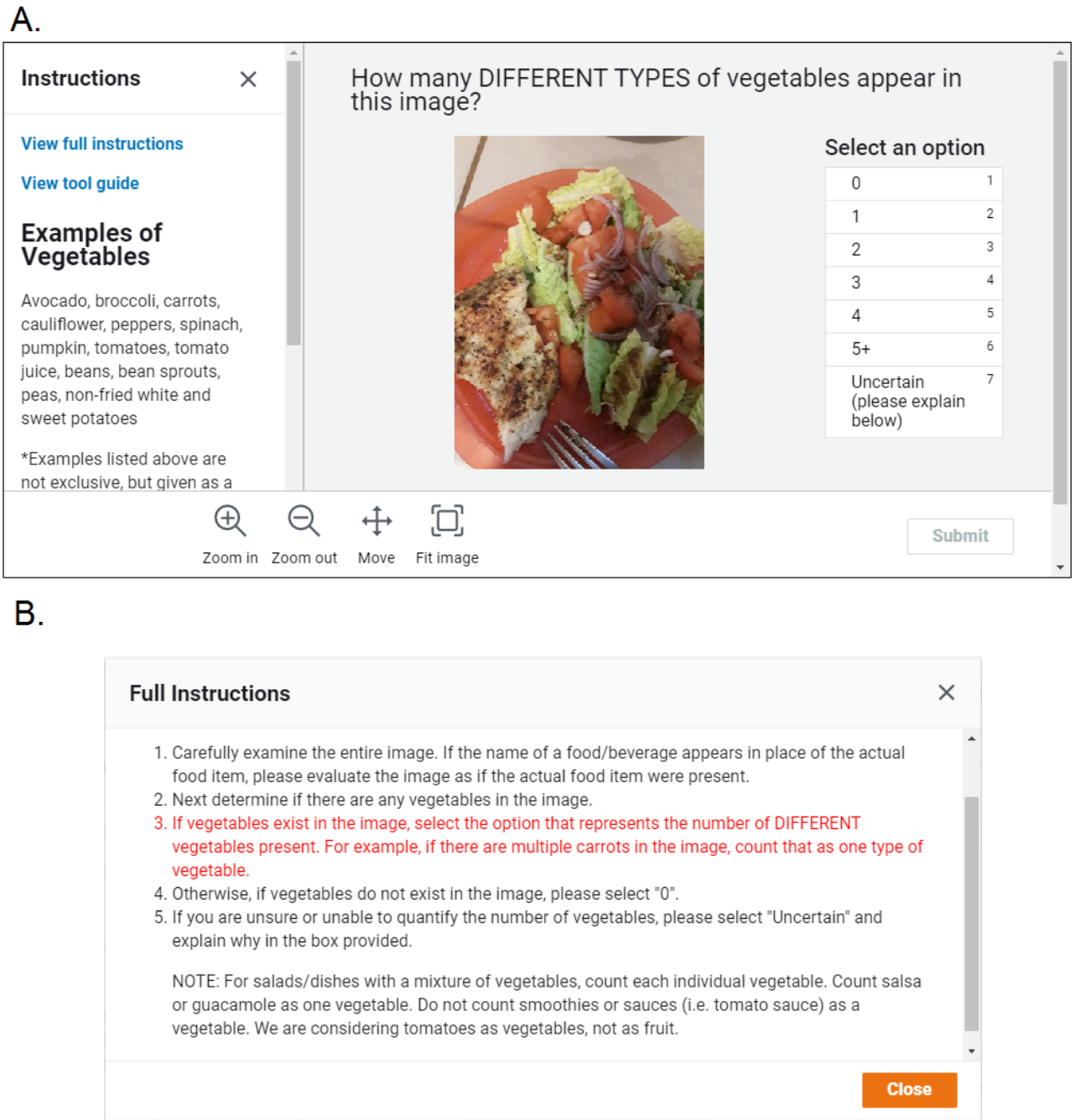
MTurk was used to process the images of participants' meals and snacks. MTurk is a crowdsourcing distributed human intelligence platform that has been used to process images for biomedical research [30-32]. Users upload discrete human intelligence tasks (HITs) that Workers can complete quickly for a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

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



XSL•FO
RenderX

**Figure 1.** Example Mechanical Turk Worker interface for vegetable tasks, including an image with examples and the full instructions.




Before uploading the images, a study team member checked each image to ensure that they did not accidentally include the participants' personal information or anything that could potentially identify participants. Each batch contained approximately 100 images (HITs), and although participants were given 1 hour to complete each HIT (ie, assign the number of food items in 1 image), on average, whole batches were completed in 1 hour and 45 minutes (SD 69 minutes). Batches can be run simultaneously; therefore, all images could be processed within the same 2-hour period. Each HIT was estimated to take the Workers 15 seconds to complete, and they were provided a US $0.05 reward upon completion of each HIT. This is equivalent to US $13 per hour, which was the minimum wage in the city of Chicago at the time of the study.

Figure 1 displays the interface that MTurk Workers were presented. Along with the prompt, "How many different types of [food category] appear in this image?", the Workers were provided with examples of the food in question and detailed instructions. Workers could select *0*, *1*, *2*, *3*, *4*, *5+*, or *uncertain* to categorize the image (Figure 1). If *uncertain* was selected, the Worker was asked to elaborate in the space provided.

## Participant Image–Captured Food Frequency

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



XSL•FO
RenderX

Images with discordant responses between Workers were evaluated by a study team member and given a final count. Weekly intake was calculated for each participant by summing the items across the images by participant ID for each food category (fruits, vegetables, salty snacks, and sweet snacks) using the final counts assigned to the images.

### Statistical Analysis

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

## Results

### Demographics

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

**Table 1.** Participant characteristics (N=48).

| Characteristics | Participants, n (%) |
|---|---|
| **Age (years)** | |
| <40 | 30 (63) |
| ≥40 | 18 (38) |
| **Race**[a] | |
| Non-Hispanic White | 23 (48) |
| Non-Hispanic Black | 5 (10) |
| Hispanic or Latina | 19 (40) |
| **Education**[a] | |
| Less than a bachelor's degree | 15 (32) |
| Bachelor's degree or higher | 32 (68) |
| **Financial burden**[a] | |
| Not very hard to pay for basics | 27 (56) |
| Hard to pay for basics | 21 (44) |

[a]One participant was missing demographic information.

### Acceptability and Data Quality

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

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



XSL•FO
RenderX

**Table 2.** Responses to survey questions on participant acceptability and data quality.

| Measure | Score, mean (SD) |
|---|---|
| **Participant acceptability**[a] | |
| Difficulty entering responses | 1.40 (0.71) |
| Difficulty understanding questions | 1.48 (0.90) |
| Difficulty uploading photographs | 2.15 (1.24) |
| **Data quality** | |
| Number of photographs taken | 21.29 (9.52) |
| Remembered to take photographs[b] | 3.85 (0.87) |
| How much did taking pictures change behavior[c] | 2.38 (1.18) |

[a]Possible options for each question in this section ranged from strongly disagree (1) to strongly agree (5).

[b]Possible options ranged from none of the time (1) to all of the time (5).

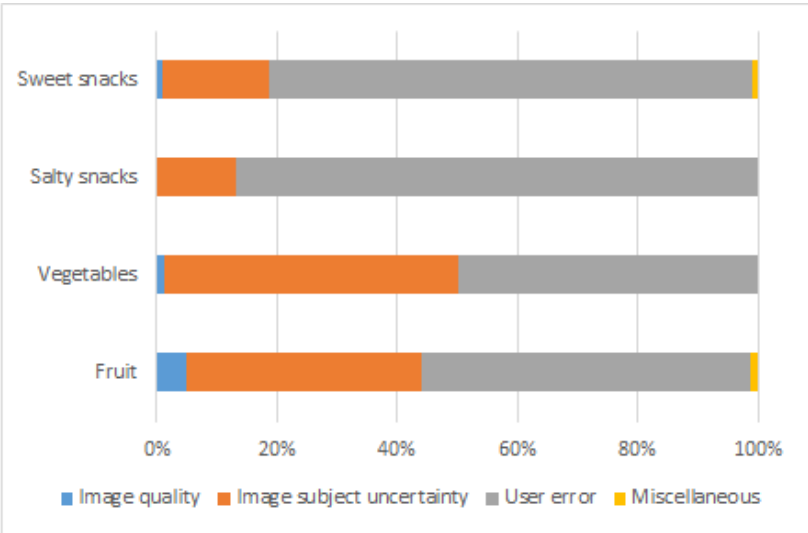
[c]Possible options ranged from not at all (1) to substantially (5).

## Feasibility of Using MTurk

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

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

**Figure 2.** Prevalence of discordant Mechanical Turk Worker response explanations by food category.






XSL•FO
RenderX

**Figure 3.** Prevalence of discordant Mechanical Turk Worker responses by types of user error.

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

### Comparison of Usual Consumption From Food Images and Dietary Screener

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

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

| Food category | Frequency, mean (SD) |
|---|---|
| **Fruit** | |
| Images | 3.58 (4.20) |
| Block Screener | 3.59 (2.32) |
| **Vegetables** | |
| Images | 9.96 (6.01) |
| Block Screener | 10.28 (4.58) |
| **Salty snacks** | |
| Images | 4.17 (3.21) |
| Block Screener | 3.85 (3.16) |
| **Sweet snacks** | |
| Images | 3.77 (4.13) |
| Block Screener | 4.51 (4.13) |

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

## Discussion

### Principal Findings

This pilot study demonstrates the feasibility of collecting data on food intake through participant-captured and crowdsource-labeled images. The method of photographing eating events in the context of an EMA study was generally



XSL•FO
RenderX

well-accepted and executed by the study participants. Most remembered to take pictures of their meals and snacks, and few reported difficulties in uploading the images. Importantly, uploading the photographs was not more difficult for participants than entering the responses. Images could be processed in a timely manner, and there was high agreement in the MTurk Worker count responses, particularly for fruits, salty snacks, and sweet snacks images, thereby supporting the feasibility of using MTurk for image classification. In addition, the weekly consumption estimated by the food images and the Screener was comparable.

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

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

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

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

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

## Limitations

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

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

## Conclusions

This pilot study demonstrates the feasibility of using participant-captured images categorized through a crowdsourcing platform to accurately depict eating events. This



XSL•FO
RenderX

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

## Acknowledgments

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

## Authors' Contributions

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

## Conflicts of Interest

None declared.

## References

1. Micha R, Kalantarian S, Wirojratana P, Byers T, Danaei G, Elmadfa I, Global Burden of Diseases, NutritionChronic Disease Expert Group. Estimating the global and regional burden of suboptimal nutrition on chronic disease: methods and inputs to the analysis. Eur J Clin Nutr 2012 Jan 14;66(1):119-129. [doi: 10.1038/ejcn.2011.147] [Medline: 21915137]
2. Micha R, Shulkin ML, Peñalvo JL, Khatibzadeh S, Singh GM, Rao M, et al. Etiologic effects and optimal intakes of foods and nutrients for risk of cardiovascular diseases and diabetes: Systematic reviews and meta-analyses from the Nutrition and Chronic Diseases Expert Group (NutriCoDE). PLoS One 2017 Apr 27;12(4):e0175149 [FREE Full text] [doi: 10.1371/journal.pone.0175149] [Medline: 28448503]
3. Willett WC, Stampfer MJ. Current evidence on healthy eating. Annu Rev Public Health 2013 Mar 18;34(1):77-95. [doi: 10.1146/annurev-publhealth-031811-124646] [Medline: 23297654]
4. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics-2020 update: a report from the American Heart Association. Circulation 2020 Mar 03;141(9):139-596 [FREE Full text] [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]
5. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May;393(10184):1958-1972. [doi: 10.1016/S0140-6736(19)30041-8]
6. Robinson E, Blissett J, Higgs S. Social influences on eating: implications for nutritional interventions. Nutr Res Rev 2013 Oct 08;26(2):166-176. [doi: 10.1017/s0954422413000127]
7. Caspi CE, Sorensen G, Subramanian S, Kawachi I. The local food environment and diet: a systematic review. Health Place 2012 Sep;18(5):1172-1187 [FREE Full text] [doi: 10.1016/j.healthplace.2012.05.006] [Medline: 22717379]
8. Hardcastle S, Thøgersen-Ntoumani C, Chatzisarantis N. Food choice and nutrition: a social psychological perspective. Nutrients 2015 Oct 21;7(10):8712-8715 [FREE Full text] [doi: 10.3390/nu7105424] [Medline: 26665419]
9. Shepherd R. Social determinants of food choice. Proc Nutr Soc 1999 Nov 28;58(4):807-812. [doi: 10.1017/s0029665199001093] [Medline: 10817147]
10. Dunton GF. Sustaining health-protective behaviors such as physical activity and healthy eating. J Am Med Assoc 2018 Aug 21;320(7):639-640 [FREE Full text] [doi: 10.1001/jama.2018.6621] [Medline: 29852046]
11. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 01;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
12. Elliston KG, Ferguson SG, Schüz B. Personal and situational predictors of everyday snacking: an application of temporal self-regulation theory. Br J Health Psychol 2017 Nov 19;22(4):854-871. [doi: 10.1111/bjhp.12259] [Medline: 28722217]
13. Schüz B, Revell S, Hills AP, Schüz N, Ferguson SG. Higher BMI is associated with stronger effects of social cues on everyday snacking behaviour. Appetite 2017 Jul 01;114:1-5. [doi: 10.1016/j.appet.2017.03.012] [Medline: 28315417]
14. Zenk SN, Horoi I, McDonald A, Corte C, Riley B, Odoms-Young AM. Ecological momentary assessment of environmental and personal factors and snack food intake in African American women. Appetite 2014 Dec;83:333-341 [FREE Full text] [doi: 10.1016/j.appet.2014.09.008] [Medline: 25239402]



XSL•FO
RenderX

15. Maugeri A, Barchitta M. A systematic review of ecological momentary assessment of diet: implications and perspectives for nutritional epidemiology. Nutrients 2019 Nov 07;11(11):2696 [FREE Full text] [doi: 10.3390/nu11112696] [Medline: 31703374]
16. Schembre SM, Liao Y, O'Connor SG, Hingle MD, Shen S, Hamoy KG, et al. Mobile ecological momentary diet assessment methods for behavioral research: systematic review. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e11170 [FREE Full text] [doi: 10.2196/11170] [Medline: 30459148]
17. Ashman A, Collins C, Brown L, Rae K, Rollo M. Validation of a smartphone image-based dietary assessment method for pregnant women. Nutrients 2017 Jan 18;9(1):73 [FREE Full text] [doi: 10.3390/nu9010073] [Medline: 28106758]
18. Boushey C, Spoden M, Delp E, Zhu F, Bosch M, Ahmad Z, et al. Reported energy intake accuracy compared to doubly labeled water and usability of the mobile food record among community dwelling adults. Nutrients 2017 Mar 22;9(3):312 [FREE Full text] [doi: 10.3390/nu9030312] [Medline: 28327502]
19. Berkman ET, Giuliani NR, Pruitt AK. Comparison of text messaging and paper-and-pencil for ecological momentary assessment of food craving and intake. Appetite 2014 Oct;81:131-137 [FREE Full text] [doi: 10.1016/j.appet.2014.06.010] [Medline: 24930596]
20. Strahler J, Nater UM. Differential effects of eating and drinking on wellbeing - an ecological ambulatory assessment study. Biol Psychol 2018 Jan;131:72-88. [doi: 10.1016/j.biopsycho.2017.01.008] [Medline: 28119068]
21. Grenard JL, Stacy AW, Shiffman S, Baraldi AN, MacKinnon DP, Lockhart G, et al. Sweetened drink and snacking cues in adolescents: a study using ecological momentary assessment. Appetite 2013 Aug;67:61-73 [FREE Full text] [doi: 10.1016/j.appet.2013.03.016] [Medline: 23583312]
22. Ashurst J, van Woerden I, Dunton G, Todd M, Ohri-Vachaspati P, Swan P, et al. The association among emotions and food choices in first-year college students using mobile-ecological momentary assessments. BMC Public Health 2018 May 02;18(1):573 [FREE Full text] [doi: 10.1186/s12889-018-5447-0] [Medline: 29716572]
23. Bejarano CM, Cushing CC. Dietary motivation and hedonic hunger predict palatable food consumption: an intensive longitudinal study of adolescents. Ann Behav Med 2018 Aug 16;52(9):773-786. [doi: 10.1093/abm/kax051] [Medline: 30124763]
24. Berge JM, Tate A, Trofholz A, Fertig A, Crow S, Neumark-Sztainer D, et al. Examining within- and across-day relationships between transient and chronic stress and parent food-related parenting practices in a racially/ethnically diverse and immigrant population : Stress types and food-related parenting practices. Int J Behav Nutr Phys Act 2018 Jan 16;15(1):7 [FREE Full text] [doi: 10.1186/s12966-017-0629-1] [Medline: 29338753]
25. Ram N, Brinberg M, Pincus AL, Conroy DE. The questionable ecological validity of ecological momentary assessment: considerations for design and analysis. Res Hum Dev 2017 Aug 10;14(3):253-270 [FREE Full text] [doi: 10.1080/15427609.2017.1340052] [Medline: 30613195]
26. Afshinnekoo E, Ahsanuddin S, Mason CE. Globalizing and crowdsourcing biomedical research. Br Med Bull 2016 Dec 23;120(1):27-33. [doi: 10.1093/bmb/ldw044] [Medline: 27941038]
27. Kershaw KN, Klikuszowian E, Schrader L, Siddique J, Van Horn L, Womack VY, et al. Assessment of the influence of food attributes on meal choice selection by socioeconomic status and race/ethnicity among women living in Chicago, USA: a discrete choice experiment. Appetite 2019 Aug 01;139:19-25 [FREE Full text] [doi: 10.1016/j.appet.2019.04.003] [Medline: 30974181]
28. Doherty TJ. The influence of aging and sex on skeletal muscle mass and strength. Curr Opin Clin Nutr Metab Care 2001 Nov;4(6):503-508. [doi: 10.1097/00075197-200111000-00007] [Medline: 11706284]
29. Lalonde I, Graham M, Slovinec-D'Angelo M, Beaton L, Brown J, Block T. Validation of the block fat/sugar/fruit/vegetable Screener in a cardiac rehabilitation setting. J Cardiopulm Rehabil Prev 2008;28(5):340. [doi: 10.1097/01.HCR.0000336162.26720.95]
30. Noronha J, Hysen E, Zhang H, Gajos KZ. Platemate: crowdsourcing nutritional analysis from food photographs. In: Proceedings of the 24th Annual Acm Symposium on User Interface Software and Technology. 2011 Presented at: UIST '11: The 24th Annual ACM Symposium on User Interface Software and Technology; October 16 - 19, 2011; Santa Barbara California USA p. 1-12. [doi: 10.1145/2047196.2047198]
31. Brady CJ, Villanti AC, Pearson JL, Kirchner TR, Gupta OP, Shah CP. Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. J Med Internet Res 2014 Oct 30;16(10):e233 [FREE Full text] [doi: 10.2196/jmir.3807] [Medline: 25356929]
32. Maier-Hein L, Mersmann S, Kondermann D, Stock C, Kenngott HG, Sanchez A, et al. Crowdsourcing for reference correspondence generation in endoscopic images. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. 2014 Presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; September 14-18, 2014; Boston, MA, USA p. 349-356. [doi: 10.1007/978-3-319-10470-6_44]
33. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]




XSL•FO
RenderX

34. Thomas JG, Doshi S, Crosby RD, Lowe MR. Ecological momentary assessment of obesogenic eating behavior: combining person-specific and environmental predictors. Obesity (Silver Spring) 2011 Aug 06;19(8):1574-1579 [FREE Full text] [doi: 10.1038/oby.2010.335] [Medline: 21273995]
35. Ghosh Roy P, Jones KK, Martyn-Nemeth P, Zenk SN. Contextual correlates of energy-dense snack food and sweetened beverage intake across the day in African American women: an application of ecological momentary assessment. Appetite 2019 Jan 01;132:73-81. [doi: 10.1016/j.appet.2018.09.018] [Medline: 30261234]
36. Six BL, Schap TE, Zhu FM, Mariappan A, Bosch M, Delp EJ, et al. Evidence-based development of a mobile telephone food record. J Am Diet Assoc 2010 Jan;110(1):74-79 [FREE Full text] [doi: 10.1016/j.jada.2009.10.010] [Medline: 20102830]
37. Vogels EA. Digital divide persists even as Americans with lower incomes make gains in tech adoption. Pew Research Center. 2021. URL: https://www.pewresearch.org/fact-tank/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/ [accessed 2021-10-26]
38. Seto E, Hua J, Wu L, Shia V, Eom S, Wang M, et al. Models of individual dietary behavior based on smartphone data: the influence of routine, physical activity, emotion, and food environment. PLoS One 2016 Apr 6;11(4):e0153085 [FREE Full text] [doi: 10.1371/journal.pone.0153085] [Medline: 27049852]

## Abbreviations

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

*Edited by G Eysenbach; submitted 27.01.21; peer-reviewed by S Wang, A Ponnada, S Victor; comments to author 29.04.21; revised version received 23.06.21; accepted 01.08.21; published 02.12.21.*

*Please cite as:*
*Harrington K, Zenk SN, Van Horn L, Giurini L, Mahakala N, Kershaw KN*
*The Use of Food Images and Crowdsourcing to Capture Real-time Eating Behaviors: Acceptability and Usability Study*
*JMIR Form Res 2021;5(12):e27512*
*URL: https://formative.jmir.org/2021/12/e27512*
*doi:10.2196/27512*
*PMID:34860666*

©Katharine Harrington, Shannon N Zenk, Linda Van Horn, Lauren Giurini, Nithya Mahakala, Kiarri N Kershaw. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Users' Perceptions Toward mHealth Technologies for Health and Well-being Monitoring in Pregnancy Care: Qualitative Interview Study

Jane Li[1*], PhD; David Silvera-Tawil[1*], PhD; Marlien Varnfield[2*], PhD; M Sazzad Hussain[1*], PhD; Vanitha Math[3*], MD

[1]Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Marsfield, Australia

[2]Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Herston, Australia

[3]Department of Obstetrics and Gynaecology, Gold Coast University Hospital, Gold Coast, Australia

[*]all authors contributed equally

**Corresponding Author:**
Jane Li, PhD
Australian e-Health Research Centre
Commonwealth Scientific and Industrial Research Organisation
Cnr Vimiera and Pembroke Roads
Marsfield, 2122
Australia
Phone: 61 02 93724163
Email: jane.li@csiro.au

## Abstract

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

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

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

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

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

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



XSL•FO
RenderX

**KEYWORDS**

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

# Introduction

## Background

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

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

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

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

## Challenges in Pregnancy mHealth Care

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

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

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

## Objective

As an initial step to inform the design of appropriate technologies to support the monitoring of health and well-being during pregnancy, we aim to conduct a qualitative study with clinicians, pregnant women, and postpartum women to understand their existing risk assessment and monitoring practices as well as their needs, interests, and preferences in mHealth. We also aim to explore the potential factors that may



XSL•FO
RenderX

influence their engagement with mHealth data collection and monitoring as part of pregnancy care. The study, conducted in the Australian context, contributes to the broader understanding of factors that motivate the use of mHealth in pregnancy care and how novel technologies can be designed and introduced to improve user engagement and long-term health monitoring during pregnancy.

# Methods

## Overview

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

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

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

## Participants

### *Female Participants*

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

### *Clinician Participants*

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

## Study Procedure

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

## Data Analysis

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



XSL•FO
RenderX

mHealth technologies, and perceived benefits and challenges in incorporating mHealth technologies in daily life and practices were identified.

## Results

### Participant Characteristics

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

**Table 1.** Characteristics of female participants (N=8).

| Characteristics | Values, n (%) |
|---|---|
| **Pregnancy stage** | |
| Second trimester | 2 (25) |
| Third trimester | 2 (25) |
| Postpartum (2-4 months) | 4 (50) |
| **Medical condition during pregnancy** | |
| Gestational diabetes | 2 (25) |
| Preeclampsia | 1 (12) |
| Iron deficiency | 1 (12) |
| No medical conditions | 4 (50) |
| **Working time during pregnancy** | |
| Full-time | 7 (88) |
| Part-time | 1 (12) |
| **First time mother** | |
| Yes | 7 (88) |
| No | 1 (12) |
| **Public or private service used** | |
| Public | 3 (38) |
| Private | 4 (50) |
| Mixed | 1 (12) |
| **Have used smartphones and apps** | |
| Yes | 8 (100) |
| No | 0 (0) |



XSL•FO
RenderX

**Table 2.** Characteristics of clinician participants (N=13).

| Characteristics | Values, n (%) |
|---|---|
| **Roles** | |
| Obstetricians | 2 (15) |
| Midwife | 3 (23) |
| Midwife, manager | 2 (15) |
| Midwife, clinical consultant | 1 (8) |
| Midwife, general practitioner liaison | 1 (8) |
| Health educator | 1 (8) |
| Social worker | 1 (8) |
| Physiotherapist | 1 (8) |
| Dietitian | 1 (8) |
| **Experience in their fields (years)** | |
| <10 | 2 (15) |
| 10-20 | 6 (46) |
| 20-30 | 4 (31) |
| >30 | 1 (8) |
| **Gender** | |
| Female | 11 (85) |
| Male | 2 (15) |

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

## Risk Assessment and Monitoring During Pregnancy

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

### *Risk Assessment and Care Needs*

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

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

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

> *It really depends on the health and well-being of the women, and sometimes I can downgrade women and sometimes I have to upgrade them as their pregnancy goes along. So sometimes, women at the beginning*



XSL•FO
RenderX

> *need quite a lot of intensive support and coordination, but once we have got that underway and we are on the right track, then it can be stepped down.* [C13]

### *Supporting Pregnant Women's Healthy Lifestyle and Well-being*

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

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

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

## Female Participants' Perception on Health Monitoring and mHealth

### *Overview*

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

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

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

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

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

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

### *Monitoring Weight*

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

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

### *Monitoring Diet*

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

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



XSL•FO
RenderX

### Monitoring Blood Pressure

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

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

### Monitoring Blood Glucose Level

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

### Monitoring Mental Health

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

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

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

### Using Mobile Apps

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

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

### Female Participants' Attitude Toward mHealth

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

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

## Clinician Participants' Perception on mHealth

### Overview

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

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

### Broad Requirements

Data collection from multiple sources, incorporating accurate information to pregnant women, clinician portal to access data,



XSL•FO
RenderX

providing features for alertness, and ease of use were among the desired features for the clinicians.

#### Health Monitoring Data

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

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

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

#### Lifestyle Data

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

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

#### Questionnaire Data

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

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

#### Incorporating Accurate Information and Feedback

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

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

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

#### Access Data by Clinicians

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

#### Alerts

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

#### Ease of Use

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

### *Clinical Considerations for mHealth Monitoring*

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

#### Gestational Diabetes

Pregnant women diagnosed with GDM receive ongoing support from diabetes educators, diabetes dietitians, and endocrine specialists from the pregnancy to postnatal stages. According to our clinician participants, women with GDM were advised



XSL•FO
RenderX

to test their glucose levels at home 4 times a day using a glucose monitor, record the results, and have periodic follow-up visits at the hospital. Existing glucose monitoring practices require extensive effort from both female patients and clinicians. Clinicians could benefit from easy access to data through web-based or mobile solutions. Our clinician participants indicated that GDM is the most common and suitable medical condition to consider for mHealth interventions.

#### Hypertension

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

#### Obesity or High BMI

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

#### Mental Health

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

#### Stillbirth Prevention

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

#### Other

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

### *Challenges*

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

#### Women's Engagement

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

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

#### Technology

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

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

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

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

#### Change of Practices

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

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

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

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



XSL•FO
RenderX

### Demonstration of Impact

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

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

## Discussion

### Potential Interests

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

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

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

### Improving Engagement With mHealth Technology

#### *Overview*

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

#### *Design Considerations*

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

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

#### *Implementation Considerations*

Our study suggests some strategies on how to work best with less motivated pregnant women. First, one possible solution to



XSL•FO
RenderX

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

## Introducing mHealth Technology in Practices

### *Overview*

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

### *Design Considerations*

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

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

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

### *Implementation Considerations*

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

## Limitations

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

## Conclusions

We have explored the aspects of current risk assessment practices, users' motivations, and concerns as well as clinical and technical factors that need to be considered when designing and introducing mHealth monitoring solutions for pregnant care. Adequate high-quality data collected through longitudinal monitoring is required for the intelligent detection of risks. We discussed technology solutions and implementation strategies to improve pregnant women's engagement with technology and data collection, which are critical for mHealth solutions to facilitate the tracking of health and behavior changes during pregnancy. Future research will include feasibility studies to inform the development of mHealth technology and evidence-based evaluation studies to understand the efficacy of mHealth solutions in supporting pregnancy care.



XSL•FO
RenderX

## Acknowledgments

This project was funded by the Commonwealth Scientific and Industrial Research Organization Precision Health Future Science Platform. The authors thank the clinicians at the Gold Coast University Hospital for their support in this study.

## Conflicts of Interest

None declared.

## References

1. Kautzky-Willer A, Bancher-Todesca D, Birnbacher R. Gestational diabetes mellitus. Acta Med Austriaca 2004;31(5):182-184. [Medline: 15747996]
2. Vatten LJ, Skjaerven R. Is pre-eclampsia more than one disease? BJOG 2004 Apr;111(4):298-302. [doi: 10.1111/j.1471-0528.2004.00071.x] [Medline: 15008762]
3. Luoto R, Mottola M, Hilakivi-Clarke L. Pregnancy and lifestyle: short- and long-term effects on mother's and her children's health. J Pregnancy 2013;2013:537526 [FREE Full text] [doi: 10.1155/2013/537526] [Medline: 23781339]
4. Oteng-Ntim E, Varma R, Croker H, Poston L, Doyle P. Lifestyle interventions for overweight and obese pregnant women to improve pregnancy outcome: systematic review and meta-analysis. BMC Med 2012 May 10;10:47 [FREE Full text] [doi: 10.1186/1741-7015-10-47] [Medline: 22574949]
5. National midwifery guidelines for consultation and referral. Australian College of Midwives. 2014. URL: https://www.midwives.org.au/resources/national-midwifery-guidelines-consultation-and-referral-3rd-edition-issue-2-2014 [accessed 2021-11-11]
6. English S, Steele A, Williams A, Blacklay J, Sorinola O, Wernisch L, et al. Modelling of psychosocial and lifestyle predictors of peripartum depressive symptoms associated with distinct risk trajectories: a prospective cohort study. Sci Rep 2018 Aug 24;8(1):12799 [FREE Full text] [doi: 10.1038/s41598-018-30874-z] [Medline: 30143660]
7. Abell SK, Shorakae S, Boyle JA, De Courten B, Stepto NK, Teede HJ, et al. Role of serum biomarkers to optimise a validated clinical risk prediction tool for gestational diabetes. Aust N Z J Obstet Gynaecol 2019 Apr;59(2):251-257. [doi: 10.1111/ajo.12833] [Medline: 29900538]
8. Moreira MW, Rodrigues JJ, Kumar N, Saleem K, Illin IV. Postpartum depression prediction through pregnancy data analysis for emotion-aware smart systems. Inf Fusion 2019 May;47:23-31. [doi: 10.1016/j.inffus.2018.07.001]
9. Velikova M, Lucas P, Spaanderman M. A predictive Bayesian network model for home management of preeclampsia. In: Artificial Intelligence in Medicine. Berlin, Heidelberg: Springer; 2011:179-183.
10. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011.
11. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
12. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
13. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 09;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
14. Boll S, Heuten W, Meyer J. From tracking to personal health. Interactions. 2016. URL: https://interactions.acm.org/archive/view/january-february-2016/from-tracking-to-personal-health [accessed 2021-11-11]
15. Huang W, Li J, Alem L. Towards preventative healthcare: a review of wearable and mobile applications. Stud Health Technol Inform 2018;251:11-14. [Medline: 29968589]
16. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
17. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
18. Silvera-Tawil D, Hussain MS, Li J. Emerging technologies for precision health: an insight into sensing technologies for health and wellbeing. Smart Health 2020 Mar;15:100100. [doi: 10.1016/j.smhl.2019.100100]
19. Rivera-Romero O, Olmo A, Muñoz R, Stiefel P, Miranda ML, Beltrán LM. Mobile health solutions for hypertensive disorders in pregnancy: scoping literature review. JMIR Mhealth Uhealth 2018 May 30;6(5):e130 [FREE Full text] [doi: 10.2196/mhealth.9671] [Medline: 29848473]
20. Chan KL, Chen M. Effects of social media and mobile health apps on pregnancy care: meta-analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11836 [FREE Full text] [doi: 10.2196/11836] [Medline: 30698533]
21. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology-based lifestyle and medical intervention apps supporting health care during pregnancy:



XSL•FO
RenderX

systematic review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]
22. Payakachat N, Rhoads S, McCoy H, Dajani N, Eswaran H, Lowery C. Using mHealth in postpartum women with pre-eclampsia: lessons learned from a qualitative study. Int J Gynaecol Obstet 2020 Jun;149(3):339-346 [FREE Full text] [doi: 10.1002/ijgo.13134] [Medline: 32119129]
23. Krishnamurti T, Davis AL, Wong-Parodi G, Fischhoff B, Sadovsky Y, Simhan HN. Development and testing of the MyHealthyPregnancy app: a behavioral decision research-based tool for assessing and communicating pregnancy risk. JMIR Mhealth Uhealth 2017 Apr 10;5(4):e42 [FREE Full text] [doi: 10.2196/mhealth.7036] [Medline: 28396302]
24. Willcox JC, van der Pligt P, Ball K, Wilkinson SA, Lappas M, McCarthy EA, et al. Views of women and health professionals on mHealth lifestyle interventions in pregnancy: a qualitative investigation. JMIR Mhealth Uhealth 2015 Oct 28;3(4):e99 [FREE Full text] [doi: 10.2196/mhealth.4869] [Medline: 26510886]
25. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
26. Shorey S, Yang YY, Dennis C. A mobile health app-based postnatal educational program (Home-but not Alone): descriptive qualitative study. J Med Internet Res 2018 Apr 19;20(4):e119 [FREE Full text] [doi: 10.2196/jmir.9188] [Medline: 29674314]
27. Chaudhry B, Faust L, Chawla NV. From design to development to evaluation of a pregnancy app for low-income women in a community-based setting. In: Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services. 2019 Presented at: MobileHCI'19; Oct 1 - 4, 2019; Taipei Taiwan. [doi: 10.1145/3338286.3340118]
28. Barry M, Doherty K, Marcano BJ, Car J, Morrison C, Doherty G. mHealth for maternal mental healthveryday wisdom in ethical design. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver Colorado USA. [doi: 10.1145/3025453.3025918]
29. Penders J, Altini M, Van Hoof C, Dy E. Wearable sensors for healthier pregnancies. Proc IEEE 2015 Feb;103(2):179-191. [doi: 10.1109/jproc.2014.2387017]
30. Goetz M, Müller M, Matthies LM, Hansen J, Doster A, Szabo A, et al. Perceptions of patient engagement applications during pregnancy: a qualitative assessment of the patient's perspective. JMIR Mhealth Uhealth 2017 May 26;5(5):e73 [FREE Full text] [doi: 10.2196/mhealth.7040] [Medline: 28550005]
31. Rodbard D. Continuous glucose monitoring: a review of successes, challenges, and opportunities. Diabetes Technol Ther 2016 Feb;18 Suppl 2:S3-13 [FREE Full text] [doi: 10.1089/dia.2015.0417] [Medline: 26784127]
32. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Front Physiol 2018 Jun 28;9:743 [FREE Full text] [doi: 10.3389/fphys.2018.00743] [Medline: 30002629]
33. Runkle J, Sugg M, Boase D, Galvin SL, C Coulson C. Use of wearable sensors for pregnancy health and environmental monitoring: descriptive findings from the perspective of patients and providers. Digit Health 2019 Feb 6;5:2055207619828220 [FREE Full text] [doi: 10.1177/2055207619828220] [Medline: 30792878]
34. Gambhir SS, Ge TJ, Vermesh O, Spitler R. Toward achieving precision health. Sci Transl Med 2018 Feb 28;10(430):eaao3612 [FREE Full text] [doi: 10.1126/scitranslmed.aao3612] [Medline: 29491186]
35. Haddad SM, Souza RT, Cecatti JG. Mobile technology in health (mHealth) and antenatal care-searching for apps and available solutions: a systematic review. Int J Med Inform 2019 Jul;127:1-8. [doi: 10.1016/j.ijmedinf.2019.04.008] [Medline: 31128820]
36. Davidson L, Boland M. Towards deep phenotyping pregnancy: a systematic review on artificial intelligence and machine learning methods to improve pregnancy outcomes. Brief Bioinform 2021 Sep 02;22(5):bbaa369 [FREE Full text] [doi: 10.1093/bib/bbaa369] [Medline: 33406530]
37. Velikova M, Lucas P, Spaanderman M. e-MomCare: a personalised home-monitoring system for pregnancy disorders. In: Electronic Healthcare. Berlin, Heidelberg: Springer; 2011:267-274.
38. Rhoads SJ, Serrano CI, Lynch CE, Ounpraseuth ST, Gauss CH, Payakachat N, et al. Exploring implementation of m-Health monitoring in postpartum women with hypertension. Telemed J E Health 2017 Oct;23(10):833-841 [FREE Full text] [doi: 10.1089/tmj.2016.0272] [Medline: 28475431]
39. van Kasteren Y, Freyne J, Hussain MS. Total knee replacement and the effect of technology on cocreation for improved outcomes and delivery: qualitative multi-stakeholder study. J Med Internet Res 2018 Mar 20;20(3):e95 [FREE Full text] [doi: 10.2196/jmir.7541] [Medline: 29559424]
40. Bashi N, Hassanzadeh H, Varnfield M, Wee Y, Walters D, Karunanithi M. Multidisciplinary smartphone-based interventions to empower patients with acute coronary syndromes: qualitative study on health care providers' perspectives. JMIR Cardio 2018 Oct 31;2(2):e10183 [FREE Full text] [doi: 10.2196/10183] [Medline: 31758781]
41. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [doi: 10.1192/bjp.150.6.782] [Medline: 3651732]



XSL•FO
RenderX

42. Moore T, Arefadib N, Deery A, West S. The first thousand days: an evidence paper. The First 1000 Days: An Evidence Paper. 2017. URL: https://www.researchgate.net/publication/320057527_The_First_Thousand_Days_An_Evidence_Paper [accessed 2021-11-18]
43. Continuous glucose monitoring. Diabetes Australia. URL: https://www.diabetesaustralia.com.au/living-with-diabetes/managing-your-diabetes/diabetes-technology/continuous-glucose-monitoring/ [accessed 2021-09-20]
44. Marko KI, Ganju N, Krapf JM, Gaba ND, Brown JA, Benham JJ, et al. A mobile prenatal care app to reduce in-person visits: prospective controlled trial. JMIR Mhealth Uhealth 2019 May 01;7(5):e10520 [FREE Full text] [doi: 10.2196/10520] [Medline: 31042154]
45. Rehman H, Kamal AK, Sayani S, Morris PB, Merchant AT, Virani SS. Using mobile health (mHealth) technology in the management of diabetes mellitus, physical inactivity, and smoking. Curr Atheroscler Rep 2017 Apr;19(4):16. [doi: 10.1007/s11883-017-0650-5] [Medline: 28243807]
46. Hearn L, Miller M, Fletcher A. Online healthy lifestyle support in the perinatal period: what do women want and do they use it? Aust J Prim Health 2013;19(4):313-318. [doi: 10.1071/PY13039] [Medline: 23899373]

## Abbreviations

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

*Edited by G Eysenbach; submitted 09.03.21; peer-reviewed by T Krishnamurti, J Wei, B Chaudhry; comments to author 29.05.21; revised version received 23.09.21; accepted 15.10.21; published 02.12.21.*

*Please cite as:*
*Li J, Silvera-Tawil D, Varnfield M, Hussain MS, Math V*
*Users' Perceptions Toward mHealth Technologies for Health and Well-being Monitoring in Pregnancy Care: Qualitative Interview Study*
*JMIR Form Res 2021;5(12):e28628*
*URL: https://formative.jmir.org/2021/12/e28628*
*doi:10.2196/28628*
*PMID:34860665*

©Jane Li, David Silvera-Tawil, Marlien Varnfield, M Sazzad Hussain, Vanitha Math. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Initial Feasibility of the "Families Moving Forward Connect" Mobile Health Intervention for Caregivers of Children With Fetal Alcohol Spectrum Disorders: Mixed Method Evaluation Within a Systematic User-Centered Design Approach

Christie Lynn McGee Petrenko[1], BA, BSc, MSc, PhD; Carson Christine Kautz-Turnbull[1], BA; Alicia Rose Roth[1], BA; Jennifer Elizabeth Parr[1], BA, MSc; Cristiano Tapparello[2], BSc, MSc, PhD; Utku Demir[2], BSc, MSc, PhD; Heather Carmichael Olson[3,4], BA, MA, PhD

[1]Mt. Hope Family Center, University of Rochester, Rochester, NY, United States

[2]Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, United States

[3]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States

[4]Center for Child Health, Behavior, and Development, Seattle Children's Research Institute, Seattle, WA, United States

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

## Abstract

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

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

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

**Results:** Across beta-test trials, 84% (38/45) of caregivers and 94% (17/18) of providers installed the FMF Connect app. Technological issues were tracked in real time with updates to address problems and expand app functionalities. On use days, caregivers averaged 20 minutes using the app; most of the time was spent watching videos in Learning Modules. Caregiver engagement with the Learning Modules varied across 5 usage pattern tiers. Overall, 67% (30/45) of caregivers posted at least once in the Family Forum. Interviews were completed by 26 caregivers and 16 providers. App evaluations generally did not differ according to usage pattern tier or demographic characteristics. Globally, app users were very positive, with 2.5 times more positive- than negative-coded segments across participants. Positive evaluations emphasized the benefits of accessible information and practical utility of the app. Informational and video content were described as especially valuable to caregivers. A number of affective and social benefits of the app were identified, aligning well with the caregivers' stated motivators for app use. Negative evaluations of user experience generally emphasized technical and navigational aspects. Refinements were made on the basis of feedback during the first beta test, which were positively received during the second round. Participants offered many valuable recommendations for continuing app refinement, which is useful in improving user experience.



XSL•FO
RenderX

**Conclusions:** The results demonstrate that the FMF Connect intervention is acceptable and feasible for caregivers raising children with FASD. They will guide subsequent app refinement before large-scale randomized testing. This study used a systematic, user-centered design approach for app development and evaluation. The approach used here may illustrate a model that can broadly inform the development of mobile health and digital parenting interventions.

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

**KEYWORDS**

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

## Introduction

### Background

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

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

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

To address significant barriers to care affecting this population, we developed a mobile health (mHealth) intervention called Families Moving Forward (FMF) Connect. mHealth, or the application of smartphones or wireless technologies to improve health, has bourgeoned since the emergence of app stores in 2008 [15]. mHealth has many potential advantages, including increasing health care capacity, providing patient access to tailored and immediate support, reducing stigma in obtaining care, and improving cost-effectiveness [16]. FMF Connect is the first known mHealth app developed and tested for FASD.

The task of developing and evaluating the FMF Connect mHealth intervention is being carried out following a systematic, user-centered design approach to app development and evaluation (Figure 1). Unfortunately, deployment of this type of systematic approach has been relatively rare for mHealth interventions [17-19]. This methodology integrates user-centered design principles, which emphasize understanding users, tasks, and environments, with the process of obtaining iterative and collaborative input from users [20-22]. There are seven phases to this approach, as operationalized in Figure 1. Of course, the process is more complex than that illustrated in Figure 1. There are certainly feedback loops between phases that indicate iterative change. This study describes the model, which involves a multidisciplinary development team and engagement of key stakeholders through focus groups and beta-testing. We also include the presentation of data from the fifth phase of the model to reveal specifically how user data from beta-testing can strategically refine and enhance app design. We note that this generalized approach can be used in the broader field of mHealth development.

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

**Figure 1.** Phases of a systematic, user-centered design approach to mobile health intervention development and evaluation.

XSL•FO
RenderX

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

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

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

## FMF Connect: A Novel mHealth Intervention

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

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

## This Study

In our systematic approach, the first four phases resulted in the development of FMF Connect app prototypes. In this study we report findings from the critical fifth phase, in which we conducted two rounds of beta-testing of the FMF Connect app prototypes to test feasibility; specifically, the feasibility and acceptability of the app intervention [27]. Each round of beta-testing involved 6-week trials, followed by key informant interviews and examination of app usage analytics. We engaged key stakeholders, including caregivers of children with FASD and medical, mental health, and other providers, to elicit their



XSL•FO
RenderX

experiences and perspectives on the FMF Connect intervention. The feasibility trial was guided by the following objectives, informed by the Eldridge et al [27] conceptual framework:

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

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

## Methods

### Study Design

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

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

### Participants

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



XSL•FO
RenderX

**Table 1.** Participant characteristics (N=63).

| Characteristics | Caregiver (n=45) | Provider (n=18) |
|---|---|---|
| **Gender, n (%)** | | |
| Female | 42 (93.3) | 17 (94.4) |
| **Age (years)** | | |
| Mean (SD) | 50.41 (11.33) | 45.28 (11.60) |
| Range | 31-73 | 28-70 |
| **Average age of children with FASD[a] (n=64)** | | |
| Mean (SD) | 9.70 (3.58) | N/A[b] |
| Range | 4-17 | N/A |
| **Caregiver ethnicity, n (%)** | | |
| Hispanic or Latinx | 3 (6.7) | 2 (11.1) |
| **Caregiver race, n (%)** | | |
| African American or Black | 4 (8.9) | 0 (0.0) |
| Asian | 0 (0.0) | 1 (5.6) |
| White | 40 (88.9) | 17 (94.4) |
| Native American or Alaska native | 2 (4.4) | 0 (0.0) |
| Native Hawaiian or Pacific islander | 0 (0.0) | 1 (5.6) |
| **Education level, n (%)** | | |
| High-school diploma or GED[c] | 2 (4.4) | 0 (0.0) |
| Some college, trade school, or Associate's degree | 13 (28.9) | 0 (0.0) |
| Bachelor's degree | 12 (26.7) | 2 (11.1) |
| Master's degree or higher | 13 (28.9) | 6 (33.3) |
| Doctoral or professional degree | 5 (11.1) | 10 (55.6) |
| **Relation to child,[d] n (%)** | | |
| Biological parent | 1 (2.2) | N/A |
| Adoptive parent | 32 (71.1) | N/A |
| Foster parent | 1 (2.2) | N/A |
| Grandparent | 9 (20.0) | N/A |
| Other relative | 2 (4.4) | N/A |
| **Family income (US $), n (%)** | | |
| 15,000-24,999 | 2 (4.4) | N/A |
| 25,000-34,999 | 3 (6.7) | N/A |
| 35,000-49,999 | 4 (8.9) | N/A |
| 50,00074,999 | 7 (15.6) | N/A |
| 75,000-99,999 | 4 (8.9) | N/A |
| 100,000-149,999 | 12 (26.7) | N/A |
| 150,000 or more | 9 (20.0) | N/A |
| Prefer not to answer | 4 (8.9) | N/A |
| **Community type, n (%)** | | |
| Rural | 7 (15.6) | N/A |
| Suburban | 33 (73.3) | N/A |
| Urban | 5 (11.1) | N/A |



XSL•FO
RenderX

| Characteristics | Caregiver (n=45) | Provider (n=18) |
|---|---|---|
| **Experience with standard FMF[e] Program, n (%)[f]** | 7 (15.6) | 9 (50.0) |
| **Round of beta-testing, n (%)** | | |
| Beta test 1 | 21 (46.7) | 17 (94.4) |
| Beta test 2 | 24 (53.3) | 1 (5.6) |
| **Operating system, n (%)** | | |
| iOS | 36 (80.0) | 17 (94.4) |
| Android | 8 (17.8) | 1 (5.6) |
| **Comfort with technology[g]** | | |
| Mean (SD) | 5.73 (1.32) | 5.83 (0.92) |
| Range | 1-7 | 4-7 |

[a]Some caregivers had more than one child with fetal alcohol spectrum disorders.

[b]N/A: not applicable.

[c]GED: General Educational Development.

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

[e]FMF: Families Moving Forward.

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

[g]Comfort with technology was measured using a Likert scale ranging from 1 (low) to 7 (high).

### *Caregivers*

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

### *Providers*

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

## Procedures

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

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



XSL•FO
RenderX

Bugs and crashes were tracked, and updates were sent to address problems or expand functionalities.

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

### Data Analysis

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

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

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

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

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

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

## Results

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

#### *Overview*

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

On use days, caregivers averaged approximately 20 minutes (mean 19.63, SD 19.59 minutes) using the app. The largest amount of time was spent watching videos in the Learning Modules (45% on average). In the Family Forum, there were 54 original posts in BT1 and 45 posts in BT2. A total of 67% (30/45) of users posted at least once in the forum.

Not unexpectedly, patterns of use varied considerably among caregivers. The standard FMF Program is similar in total time spent on other parent training programs. The FMF Program typically involves 6 to 9 months of therapist-delivered content in a collaborative therapeutic relationship with caregivers



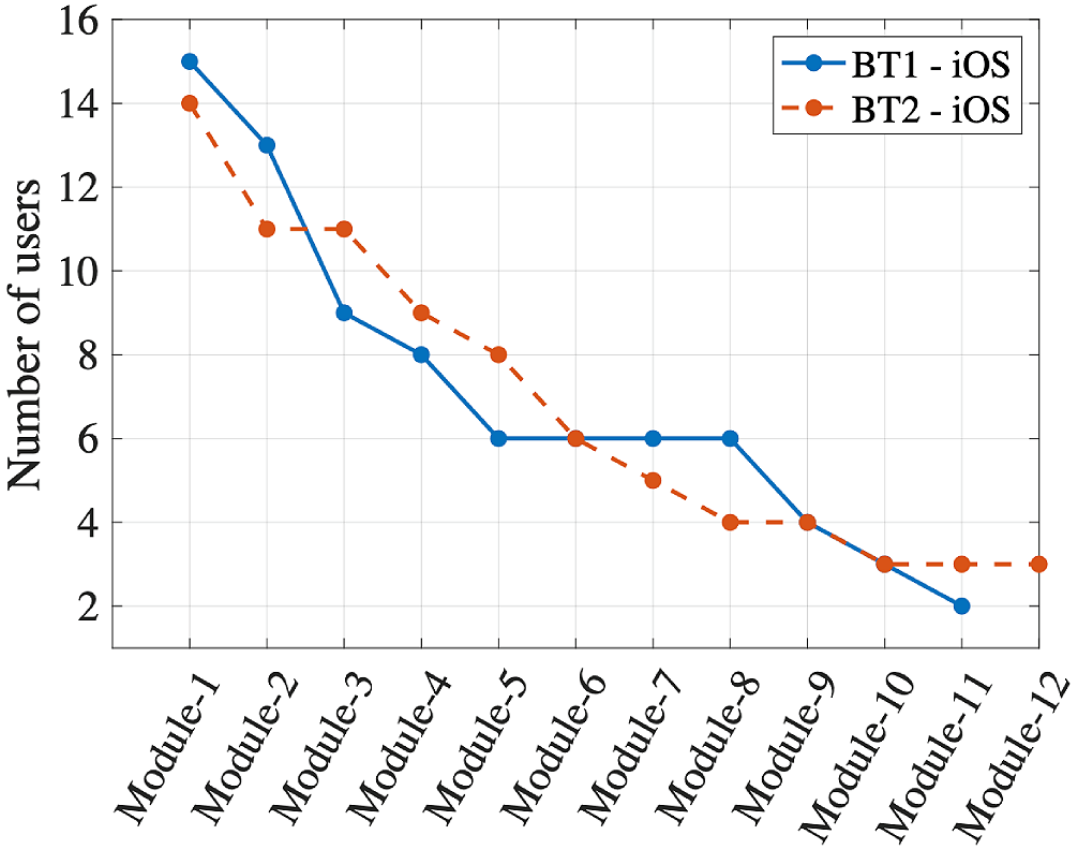
XSL•FO
RenderX

(sessions every other week). Therefore, we did not necessarily expect users to complete all 12 modules in the initial 6-week test. Usage data differed according to the operating system and will be discussed separately.

### *iOS Usage*

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

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





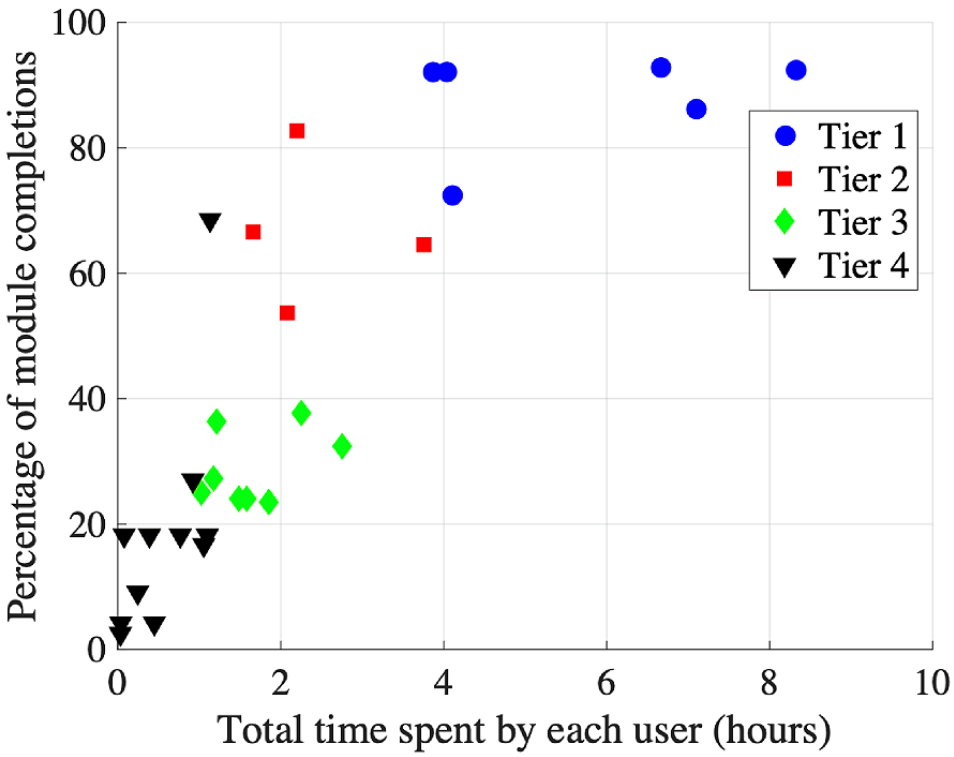

XSL•FO
RenderX

**Table 2.** Learning module usage tiers for iOS users who installed the Families Moving Forward Connect app.[a]

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

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

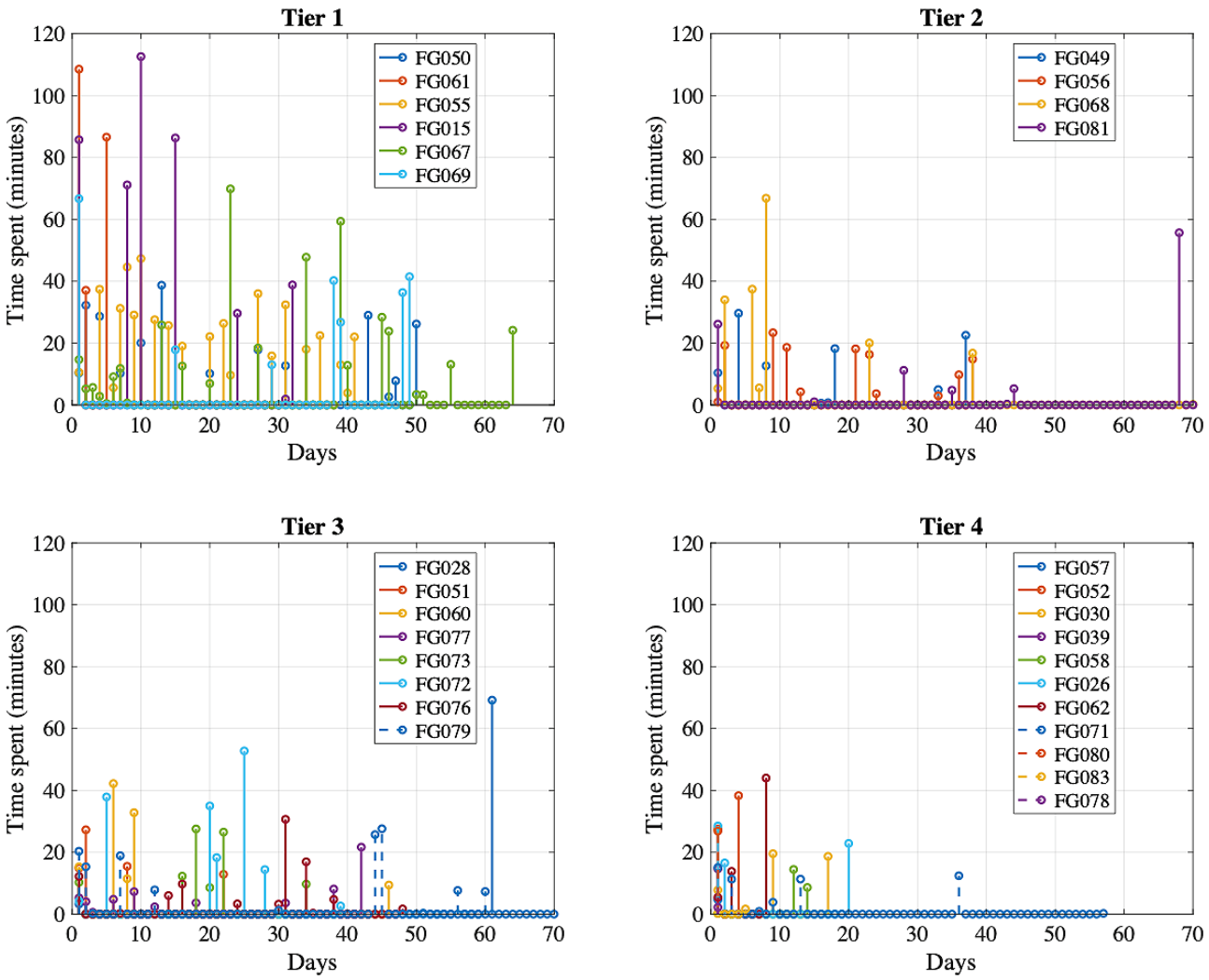
**Figure 4.** Tier classifications of usage for iOS users of the Families Moving Forward Connect mobile health intervention.






XSL•FO
RenderX

**Figure 5.** Total time spent in Learning Modules by day of use for iOS users grouped by usage tier.




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

iOS usage tier membership was also examined by participant characteristics through visual inspection and statistics (chi-square and analysis of variance). Usage tier did not differ by caregiver age ($F_3$=1.41; $P$=.26), level of education ($\chi^2_9$=6.15; $P$=.73), previous receipt of standard FMF ($\chi^2_3$=2.77, $P$=.43), target child age ($F_3$=2.46; $P$=.09), or income level ($\chi^2_{18}$=23.17, $P$=.18). Tiers varied according to comfort with technology ($F_3$=9.07, $P$<.001). Participants who rated themselves lower on comfort with technology were surprisingly more likely to be in tier 1.

### *Android Usage*

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



XSL•FO
RenderX

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

### *Overview*

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

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

### *Findings From Values Coding*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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



XSL•FO
RenderX

limited time wisely and valued choice and autonomy in self-directed learning. A provider (FG040) stated as follows:

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

A caregiver (FG051) emphasized the value of choice:

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

### *Evaluative Coding: Positives*

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

#### Global Impressions

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

> *This makes it easy for me, it's right there at my fingertips.*

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

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

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

#### Learning Modules and Libraries

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

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

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

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

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

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

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

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

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

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

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

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

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

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



XSL•FO
RenderX

> *I've been in the trenches for many years (laughs). It's nothing is new to me. And I know-like, I guess I kept thinking as I was using it, this would be super helpful for someone who was in my shoes 8 years ago.*

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

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

#### Family Forum

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

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

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

#### Dashboard and Notebook

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

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

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

#### Motivators and Facilitators of App Use

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

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

Another caregiver (FG068) shared as follows:

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

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

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

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

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

### *Evaluative Coding: Negatives*

#### Overview

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

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

#### Technological Problems

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

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



XSL•FO
RenderX

> *In this app, it was just stuck, I couldn't get out of it, and I couldn't do anything unless I completely closed it down.*

A provider (FG036) reported problems with video loading:

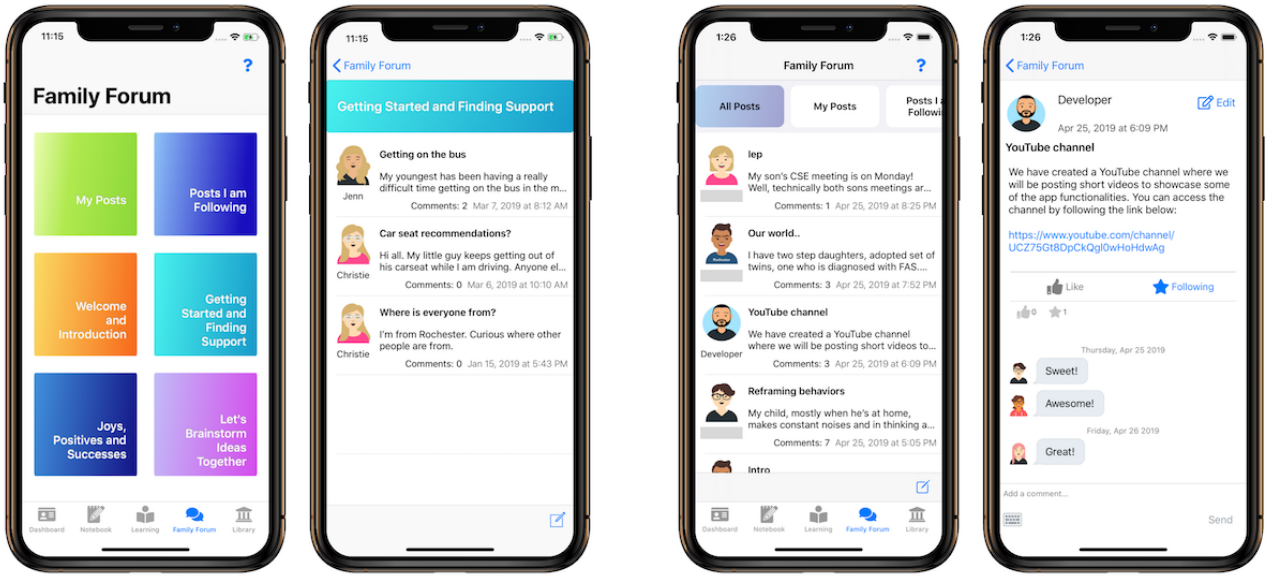
> *I could not get the first video to play. Umm, and that was something that had happened throughout, like me trying to watch the videos is they just keep buffering and buffering.*

### Navigation: Family Forum

On the basis of previous feedback regarding the design of the FMF Connect app [26], the Family Forum was initially laid out to allow gradual access to subforums tied to Learning Module completion to support privacy and shared knowledge. However, participants in BT1 did not find the Family Forum interface to be very intuitive (Figure 6A). A caregiver (FG050) explained as follows:

> *I found that there's too many boxes. There's too many sections and to go check on each one and to see what people wrote and see what their comments are as opposed to Facebook's way, like everything is there.*

**Figure 6.** Families Moving Forward Connect interface design changes for the Family Forum during (A) round 1 of beta-testing (BT1) to (B) round 2 of beta-testing (BT2).




(A) (B)

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

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

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

### Navigation: Learning Modules and Library

As mentioned previously, the step-by-step progression of the Learning Modules received a mixed evaluation. Reactions to move through the content in order varied by usage tier and previous experience with FASD. Some experienced users, with robust tier 1 use, identified that content was redundant for them. However, none of these experienced users described their experience as tedious or found that step-by-step progression was a barrier. These themes were only present for lower-use tiers 2 to 4. Negative evaluations related to step-by-step progression occurred more often for participants with a greater degree of previous knowledge and experience with FASD. One caregiver (FG066) expressed this sentiment as follows:

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

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

> *Maybe it's because I already came in with a fair amount of knowledge, but I…wasn't a big fan of different things opening up as I went. I would have preferred to have jumped in and seen everything immediately.*



XSL•FO
RenderX

Participants in BT1 found the number of videos and screens per module to be overwhelming and a barrier to completion. For example, one caregiver (FG051) explained as follows:

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

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

#### Barriers

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

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

Another caregiver (FG071) also commented as follows:

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

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

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

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

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

### *Evaluative Coding: Recommendations*

The participants offered a number of useful recommendations to improve navigation and enhance engagement with the app (Table 3). Some of these recommendations were directly related to aspects that were negatively evaluated. As mentioned above, these led to refinements to the app in between trials, and additional changes are underway.



XSL•FO
RenderX

**Table 3.** Recommendations offered by participants for further refinement of the Families Moving Forward Connect mobile health intervention.

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

[a]FASD: fetal alcohol spectrum disorders.

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

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

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

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

Caregivers also value resources to help them better understand their children. Most caregivers liked the idea of a behavior



XSL•FO
RenderX

tracker to help them monitor and notice patterns in complex behavior problems.

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

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

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

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

Another caregiver (FG069) also noted as follows:

> *You know what I'd really like to see is, is this program targeted at medical providers!*

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

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

Similarly, a caregiver (FG042) stated as follows:

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

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

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

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

Another caregiver (FG043) described the following:

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

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

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

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

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

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

## Discussion

### Principal Findings

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



XSL•FO
RenderX

Two primary research objectives were examined to evaluate the feasibility of the FMF Connect intervention during this phase of our systematic approach.

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

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

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

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

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

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



XSL•FO
RenderX

**Table 4.** Comparison of design features of the Families Moving Forward (FMF) Connect mobile health intervention with thematic checklist of features to improve user engagement.

| Design feature themes that improve engagement Wei et al [37] | Existing features in FMF Connect beta tests | Recommended features for future development |
|---|---|---|
| Interface esthetics | • Pleasing color scheme<br>• Positive evaluation of graphics | __[a] |
| Navigation | • Easy to use (iOS)<br>• Table of contents (BT2[b]) | • Tutorials/overlays<br>• Direct search feature<br>• Navigation shortcuts |
| Personalization | • Selected videos (BT2)<br>• Profile graph<br>• Exercises—provision of goal setting and feedback | • Open access to content<br>• Multiple children<br>• Personalize notifications<br>• Behavior tracker tool |
| Reinforcement | • Messages of congratulations<br>• Weekly emails<br>• Tip of the day (BT2) | • Notifications<br>• Badges |
| Communication | • Family Forum | • Coaching<br>• State or regional forums |
| Message presentation | • Simple language<br>• Positive and nonstigmatizing tone<br>• Videos<br>• Pictures<br>• Font sizes and colors to highlight information<br>• Checks for understanding in games/quizzes | • Closed captioning<br>• Spanish language |
| Credibility | • Evidence-based information from credible source<br>• Encrypted and password-protected | — |

[a]None provided by participants.

[b]BT2: second round of beta-testing.

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

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

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

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



XSL•FO
RenderX

additional barriers that could impact caregiver use of the FMF Connect on the basis of their experiences serving diverse families, including literacy level, English as a second language, caregiver age, and comfort with technology.

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

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

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

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

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

### Strengths and Limitations

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

Study findings are limited by the perspectives of participants sampled. As is true in many studies, all participants were volunteers, contributing to the possibility of selection bias. The current sample size is relatively large for beta-testing feasibility studies and is considered sufficient for the primary objectives of this study. However, it is possible that valuable perspectives may have been missed. Although the consistency of themes was very high across both caregivers and providers, some demographics of this diverse clinical population are less well represented. For example, only 7% of the parents and 6% of the providers were men. Overrepresentation of women is common in studies involving primary caregivers [53,54]. The sample represented primarily adoptive parents, although the perspectives of relative caregivers and biological mothers are represented. Racial and ethnic diversity is also somewhat lower than in the general population, and all participants were English-speaking (although some may have had fluency in other languages). The study was also limited to participants who were able to afford smartphones, WiFi, or data plans. Participant income spanned all queried levels but was still skewed above average relative to national statistics. It is notable that the inclusion of provider perspectives provided additional insights into potential barriers



XSL•FO
RenderX

and recommendations for families that may not have been represented in the sample.

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

## Conclusions and Future Directions

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

## Acknowledgments

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

## Authors' Contributions

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

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

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

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured questioning route for the focus group and individual interviews.
[DOCX File , 16 KB - formative_v5i12e29687_app1.docx ]



XSL•FO
RenderX

## References


1. Hoyme HE, Kalberg WO, Elliott AJ, Blankenship J, Buckley D, Marais A, et al. Updated clinical guidelines for diagnosing fetal alcohol spectrum disorders. Pediatrics 2016 Aug;138(2):e20154256 [FREE Full text] [doi: 10.1542/peds.2015-4256] [Medline: 27464676]
2. Mattson SN, Bernes GA, Doyle LR. Fetal alcohol spectrum disorders: a review of the neurobehavioral deficits associated with prenatal alcohol exposure. Alcohol Clin Exp Res 2019 Jun;43(6):1046-1062 [FREE Full text] [doi: 10.1111/acer.14040] [Medline: 30964197]
3. Popova S, Lange S, Shield K, Mihic A, Chudley AE, Mukherjee RA, et al. Comorbidity of fetal alcohol spectrum disorder: a systematic review and meta-analysis. Lancet 2016 Mar 05;387(10022):978-987. [doi: 10.1016/S0140-6736(15)01345-8] [Medline: 26777270]
4. May PA, Chambers CD, Kalberg WO, Zellner J, Feldman H, Buckley D, et al. Prevalence of fetal alcohol spectrum disorders in 4 US communities. JAMA 2018 Feb 06;319(5):474-482 [FREE Full text] [doi: 10.1001/jama.2017.21896] [Medline: 29411031]
5. Chasnoff IJ, Wells AM, King L. Misdiagnosis and missed diagnoses in foster and adopted children with prenatal alcohol exposure. Pediatrics 2015 Feb;135(2):264-270. [doi: 10.1542/peds.2014-2171] [Medline: 25583914]
6. Petrenko CL, Tahir N, Mahoney EC, Chin NP. Prevention of secondary conditions in fetal alcohol spectrum disorders: identification of systems-level barriers. Matern Child Health J 2014 Aug;18(6):1496-1505 [FREE Full text] [doi: 10.1007/s10995-013-1390-y] [Medline: 24178158]
7. Ryan DM, Bonnett DM, Gass CB. Sobering thoughts: town Hall meetings on fetal alcohol spectrum disorders. Am J Public Health 2006 Dec;96(12):2098-2101. [doi: 10.2105/AJPH.2005.062729] [Medline: 17077397]
8. Burles M, Holtslander L, Bocking S, Brenna B. Strengths and challenges: a young adult pictures FASD through photovoice. Rev Disabl Stud Int J 2018 Mar 18;14(1):1-20 [FREE Full text]
9. Knorr L, McIntyre LJ. Resilience in the face of adversity: stories from adults with fetal alcohol spectrum disorders. Except Educ Int 2016;26(1):53-75. [doi: 10.5206/eei.v26i1.7735]
10. Mukherjee R, Wray E, Commers M, Hollins S, Curfs L. The impact of raising a child with FASD upon carers: findings from a mixed methodology study in the UK. Adopt Foster 2013 Apr 18;37(1):43-56. [doi: 10.1177/0308575913477331]
11. Petrenko CL, Alto ME, Hart AR, Freeze SM, Cole LL. "I'm Doing My Part, I Just Need Help From the Community": intervention implications of foster and adoptive parents' experiences raising children and young adults with FASD. J Fam Nurs 2019 May;25(2):314-347 [FREE Full text] [doi: 10.1177/1074840719847185] [Medline: 31079560]
12. Sanders J, Buck G. A long journey: biological and non-biological parents' experiences raising children with FASD. J Popul Ther Clin Pharmacol 2010;17(2):e308-e322 [FREE Full text]
13. Ordenewitz LK, Weinmann T, Schlüter JA, Moder JE, Jung J, Kerber K, et al. Evidence-based interventions for children and adolescents with fetal alcohol spectrum disorders - A systematic review. Eur J Paediatr Neurol 2021 Jul;33:50-60. [doi: 10.1016/j.ejpn.2021.02.001] [Medline: 34058625]
14. Reid N, Dawe S, Shelton D, Harnett P, Warner J, Armstrong E, et al. Systematic review of fetal alcohol spectrum disorder interventions across the life span. Alcohol Clin Exp Res 2015 Dec;39(12):2283-2295. [doi: 10.1111/acer.12903] [Medline: 26578111]
15. Xu W, Liu Y. mHealthApps: a repository and database of mobile health apps. JMIR Mhealth Uhealth 2015 Mar 18;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
16. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015 May 19;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
17. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
18. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
19. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
20. Baek E, Cagiltay K, Boling E, Frick T. User-centered design and development. In: Handbook of Research on Educational Communications and Technology: A Project of the Association for Educational Communications and Technology. Oxfordshire: Routledge; 2008:659-671.
21. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
22. Dopp AR, Parisi KE, Munson SA, Lyon AR. A glossary of user-centered design strategies for implementation experts. Transl Behav Med 2019 Nov 25;9(6):1057-1064. [doi: 10.1093/tbm/iby119] [Medline: 30535343]




XSL•FO
RenderX

23. Bertrand J, Interventions for Children with Fetal Alcohol Spectrum Disorders Research Consortium. Interventions for children with fetal alcohol spectrum disorders (FASDs): overview of findings for five innovative research projects. Res Dev Disabil 2009;30(5):986-1006. [doi: 10.1016/j.ridd.2009.02.003] [Medline: 19327965]
24. Olson H, Montague R. An innovative look at early intervention for children affected by prenatal alcohol exposure. In: Prenatal Alcohol Use and FASD: a Model for Diagnosis, Assessment and New Directions in Research and Multimodal Treatment. Potomac, MD: Bentham Science Publishers; 2011:64-106.
25. Families moving forward program homepage. Families Moving Forward Program. URL: http://depts.washington.edu/fmffasd/home [accessed 2021-11-07]
26. Petrenko CL, Parr J, Kautz C, Tapparello C, Olson HC. A mobile health intervention for fetal alcohol spectrum disorders (families moving forward connect): development and qualitative evaluation of design and functionalities. JMIR Mhealth Uhealth 2020 Apr 06;8(4):e14721 [FREE Full text] [doi: 10.2196/14721] [Medline: 32250274]
27. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016 Mar 15;11(3):e0150205 [FREE Full text] [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]
28. Schoonenboom J, Johnson RB. How to construct a mixed methods research design. Kolner Z Soz Sozpsychol 2017;69(Suppl 2):107-131 [FREE Full text] [doi: 10.1007/s11577-017-0454-1] [Medline: 28989188]
29. Teddlie C, Tashakkori A. A general typology of research designs featuring mixed methods. Res Schools 2006;13(1):12-28 [FREE Full text]
30. Liamputtong P. Qualitative data analysis: conceptual and practical considerations. Health Promot J Austr 2009 Aug;20(2):133-139. [doi: 10.1071/he09133] [Medline: 19642962]
31. Miles M. Qualitative Data Analysis: A Methods Sourcebook. New York: SAGE Publications; 2013.
32. Saldaña J. The Coding Manual for Qualitative Researchers. New York: SAGE Publications; 2009.
33. de Haan AM, Boon AE, de Jong JT, Hoeve M, Vermeiren RR. A meta-analytic review on treatment dropout in child and adolescent outpatient mental health care. Clin Psychol Rev 2013 Jul;33(5):698-711. [doi: 10.1016/j.cpr.2013.04.005] [Medline: 23742782]
34. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
35. Percentage of mobile apps that have been used only once from 2010 to 2019. Statista. URL: https://www.statista.com/statistics/271628/percentage-of-apps-used-once-in-the-us/ [accessed 2021-11-07]
36. Overall app benchmarks H2. Localytics. 2017. URL: https://uplandsoftware.com/localytics/resources/cheat-sheet/overall-app-benchmarks-h2-2017/ [accessed 2021-11-07]
37. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
38. Chamberlain K, Reid N, Warner J, Shelton D, Dawe S. A qualitative evaluation of caregivers' experiences, understanding and outcomes following diagnosis of FASD. Res Dev Disabil 2017 Apr;63:99-106. [doi: 10.1016/j.ridd.2016.06.007] [Medline: 27346290]
39. Watson SL, Hayes SA, Coons KD, Radford-Paz E. Autism spectrum disorder and fetal alcohol spectrum disorder. Part II: a qualitative comparison of parenting stress. J Intellect Dev Disabil 2013 Jun;38(2):105-113. [doi: 10.3109/13668250.2013.788137] [Medline: 23672659]
40. Hundert AS, Huguet A, Green CR, Hewitt AJ, Mushquash CJ, Muhajarine N, et al. Usability testing of guided internet-based parent training for challenging behavior in children with fetal alcohol spectrum disorder (strongest families FASD). J Popul Ther Clin Pharmacol 2016;23(1):e60-e76. [Medline: 27115205]
41. Turner K, Reynolds JN, McGrath P, Lingley-Pottie P, Huguet A, Hewitt A, et al. Guided internet-based parent training for challenging behavior in children with fetal alcohol spectrum disorder (Strongest Families FASD): study protocol for a randomized controlled trial. JMIR Res Protoc 2015 Oct 13;4(4):e112 [FREE Full text] [doi: 10.2196/resprot.4723] [Medline: 26462968]
42. Jackson JB, Steward SR, Roper SO, Muruthi BA. Support group value and design for parents of children with severe or profound intellectual and developmental disabilities. J Autism Dev Disord 2018 Dec;48(12):4207-4221. [doi: 10.1007/s10803-018-3665-z] [Medline: 29980899]
43. Alsem MW, Ausems F, Verhoef M, Jongmans MJ, Meily-Visser JM, Ketelaar M. Information seeking by parents of children with physical disabilities: an exploratory qualitative study. Res Dev Disabil 2017 Jan;60:125-134. [doi: 10.1016/j.ridd.2016.11.015] [Medline: 27914304]
44. Resch JA, Mireles G, Benz MR, Grenwelge C, Peterson R, Zhang D. Giving parents a voice: a qualitative study of the challenges experienced by parents of children with disabilities. Rehabil Psychol 2010 May;55(2):139-150. [doi: 10.1037/a0019473] [Medline: 20496968]
45. Jones J, Passey J. Family adaptation, coping and resources: parents of children with developmental disabilities and behaviour problems. J Dev Disabil 2004;11(1):31-46.




XSL•FO
RenderX

46. Koren GI, Fantus E, Nulman I. Managing fetal alcohol spectrum disorder in the public school system: a needs assessment pilot. Can J Clin Pharmacol 2010;17(1):e79-e89. [Medline: 20160330]
47. Mukherjee R, Wray E, Curfs L, Hollins S. Knowledge and opinions of professional groups concerning FASD in the UK. Adopt Foster 2015 Sep 28;39(3):212-224. [doi: 10.1177/0308575915598931]
48. Roozen S, Stutterheim SE, Bos AE, Kok G, Curfs LM. Understanding the social stigma of fetal alcohol spectrum disorders: from theory to interventions. Found Sci 2020 May 29. [doi: 10.1007/s10699-020-09676-y]
49. Choate P, Badry D. Stigma as a dominant discourse in fetal alcohol spectrum disorder. Adv Dual Diagnos 2018 Dec;12(3):36-52. [doi: 10.1108/add-05-2018-0005]
50. Parsons D, Cordier R, Vaz S, Lee HC. Parent-mediated intervention training delivered remotely for children with autism spectrum disorder living outside of urban areas: systematic review. J Med Internet Res 2017 Aug 14;19(8):e198 [FREE Full text] [doi: 10.2196/jmir.6651] [Medline: 28807892]
51. Skorka K, McBryde C, Copley J, Meredith PJ, Reid N. Experiences of children with fetal alcohol spectrum disorder and their families: a critical review. Alcohol Clin Exp Res 2020 Jun;44(6):1175-1188. [doi: 10.1111/acer.14335] [Medline: 32282931]
52. Eborall H, Morton K. Use of focus groups in developing behavioural mHealth interventions: a critical review. In: A New Era in Focus Group Research: Challenges, Innovation and Practice. London: Palgrave Macmillan; Jun 2017:109-127.
53. Davison KK, Gicevic S, Aftosmes-Tobio A, Ganter C, Simon CL, Newlan S, et al. Fathers' representation in observational studies on parenting and childhood obesity: a systematic review and content analysis. Am J Public Health 2016 Nov;106(11):e14-e21. [doi: 10.2105/AJPH.2016.303391] [Medline: 27631735]
54. Davison KK, Charles JN, Khandpur N, Nelson TJ. Fathers' perceived reasons for their underrepresentation in child health research and strategies to increase their involvement. Matern Child Health J 2017 Feb;21(2):267-274 [FREE Full text] [doi: 10.1007/s10995-016-2157-z] [Medline: 27473093]
55. Collaborative Initiative on Fetal Alcohol Spectrum Disorder. URL: https://www.cifasd.org/ [accessed 2021-11-16]

## Abbreviations

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

*Edited by G Eysenbach; submitted 20.04.21; peer-reviewed by B Nievas Soriano, IN Gomez; comments to author 28.06.21; revised version received 20.09.21; accepted 09.10.21; published 02.12.21.*

*Please cite as:*
*Petrenko CLM, Kautz-Turnbull CC, Roth AR, Parr JE, Tapparello C, Demir U, Olson HC*
*Initial Feasibility of the "Families Moving Forward Connect" Mobile Health Intervention for Caregivers of Children With Fetal Alcohol Spectrum Disorders: Mixed Method Evaluation Within a Systematic User-Centered Design Approach*
*JMIR Form Res 2021;5(12):e29687*
*URL: https://formative.jmir.org/2021/12/e29687*
*doi:10.2196/29687*
*PMID:34860661*

©Christie Lynn McGee Petrenko, Carson Christine Kautz-Turnbull, Alicia Rose Roth, Jennifer Elizabeth Parr, Cristiano Tapparello, Utku Demir, Heather Carmichael Olson. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Conversational Agent for Healthy Lifestyle Behavior Change: Web-Based Feasibility Study

Dhakshenya Ardhithy Dhinagaran[1], BSc(Hons); Thirunavukkarasu Sathish[2,3], MBBS, MPH, PhD; AiJia Soong[1], BSc, MPH; Yin-Leng Theng[4], PhD; James Best[1], MD, MBBS; Lorainne Tudor Car[1,5], MD, MSc, PhD

[1]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore, Singapore

[2]Population Health Research Institute, McMaster University, Hamilton, ON, Canada

[3]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[4]Centre for Healthy and Sustainable Cities, Nanyang Technological University, Singapore, Singapore, Singapore

[5]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

**Corresponding Author:**
Lorainne Tudor Car, MD, MSc, PhD
Lee Kong Chian School of Medicine
Nanyang Technological University, Singapore
Clinical Sciences Building
11 Mandalay Road
Singapore, 308232
Singapore
Phone: 65 69041258
Email: lorainne.tudor.car@ntu.edu.sg

## Abstract

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

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

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

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

**Conclusions:** The delivery of a conversational agent for healthy lifestyle behavior change through Facebook Messenger was feasible and acceptable. We were unable to recruit our planned sample solely using the free options in Facebook. However, participant retention and conversational agent engagement rates were high. Our findings provide important insights to inform the design of a future randomized controlled trial.

***(JMIR Form Res 2021;5(12):e27956)*** doi:10.2196/27956



XSL•FO
RenderX

**KEYWORDS**

chatbot; conversational agents; behavior change; healthy lifestyle behavior change; pilot study; feasibility trial; usability; acceptability; preliminary efficacy; mobile phone

# Introduction

## Background

In recent years, there has been a notable increase in the incidence of chronic disease, especially among the younger population [1]. These chronic diseases include obesity and type 2 diabetes [1]. In terms of diabetes prevalence, Singapore, with 600,000 adults living with diabetes, ranks second among high-income countries, and obesity levels are also on the rise [2]. In addition, 34% of the men and 39% of the women in Singapore do not reach the weekly target of 150 minutes of moderate-intensity activity per week, leaving people at higher risk of developing diseases such as diabetes [3]. People with diabetes live shorter lives—by at least 10 years—and have a lower quality of life (QoL) than those without diabetes [4,5]. Prediabetes, a precursor to diabetes, affected 15.5% of the Singaporean adults in 2010 (ie, 1 in 7), and this figure is estimated to increase to 24.9% by 2035 [6]. Prediabetes increases the risk of heart disease, and if untreated, over time, most people with prediabetes transition to diabetes [7]. Research has shown that lifestyle interventions (such as increasing physical activity and eating a healthy diet) delivered by trained health care professionals can help to promote healthy weight loss [8-10]. In addition, lifestyle change through high-risk and population-based approaches were also directly associated with a reduction in the incidence of type 2 diabetes, accentuating the efficacy of healthy lifestyle behavior change for diabetes prevention [8].

Achieving healthy lifestyle behavior changes independently can be challenging, and support from experts such as dietitians or exercise physiologists has been shown to be more effective [8]. However, access to experts at a population level may not be feasible or affordable. A potentially more accessible alternative to in-person support and supervision could be novel digital health interventions such as conversational agents. Conversational agents, or chatbots, are computer programs designed to mimic human-to-human conversations in the form of either text messaging or verbal discourse [11]. The heightened accessibility, personalization, and efficiency that conversational agents offer highlight the potential for conversational agents to improve patient care [11-13]. Conversational agents enable 2-way communication, and their text- or speech-based method of communication makes them suitable for a variety of target populations, ranging from young children to older people. The application of conversational agents in health care is gaining traction in a number of medical fields, including health care service provision, chronic disease management, and patient education [14]. They can be delivered through a variety of means: messaging apps, individual apps, or even standalone devices [14].

Singapore is a technologically savvy country, and citizens avidly use messaging apps. In addition, Singapore's ministry of health has proposed the increasing use of conversational agents in health care in tackling issues such as the rising chronic disease burden and the aging population [15] that can lead to more primary care appointments. The ministry envisions a near future where a conversational agent can collect a patient's history from them before their consultation, streamlining the primary care visit and thus cutting down waiting times. This makes Singapore an ideal place for the evaluation of novel mobile health interventions such as conversational agents. Moreover, health programs delivered over the internet have shown success, as exemplified by web-based interventions for smoking [16], alcohol intake [17], sexual health [18], cancer screening [19], physical activity [20], and diet modifications [21]. The ubiquity of the internet makes these programs easily accessible to a diverse group.

## Objective

The evidence for the use of conversational agents for healthy lifestyle behavior change from trials is limited. The feasibility and acceptability of implementing and evaluating the use of novel interventions are essential for informing potential future trials. Correspondingly, we aim to assess the feasibility, acceptability, and preliminary efficacy of the use of conversational agents for healthy lifestyle behavior changes in the general population in Singapore.

# Methods

## Approval and Consent

This web-based single-arm feasibility study was approved by the Nanyang Technological University Ethics Committee (IRB-2018-11-032). All participants signed their digital informed consent before embarking on the study.

## Participants

Participants were eligible if they were aged above 21 years, were Singapore citizens or permanent residents, owned a smartphone, and had a Facebook Messenger (Facebook, Inc) account. Prospective participants were excluded if they were pregnant or had any of the following conditions: cancer, chronic liver disease, chronic kidney disease, a neurodegenerative condition, heart disease, stroke, a physical disability, hypertension, or a condition that does not allow for regular physical activity. Eligibility was confirmed by having participants complete an eligibility questionnaire, after which they were asked to provide informed consent on a digital form sent to them through email.

## Recruitment

The participants were recruited on the web through Facebook in August-September 2019. A digital poster listing the study aims and eligibility criteria was uploaded on relevant Facebook pages focused on healthy living, such as *Singapore fitness and health community* and *Singapore healthy cooking*. In addition, we used snowballing in our recruitment; therefore, participants were also procured through redistribution of our study poster through messaging apps or through word of mouth.



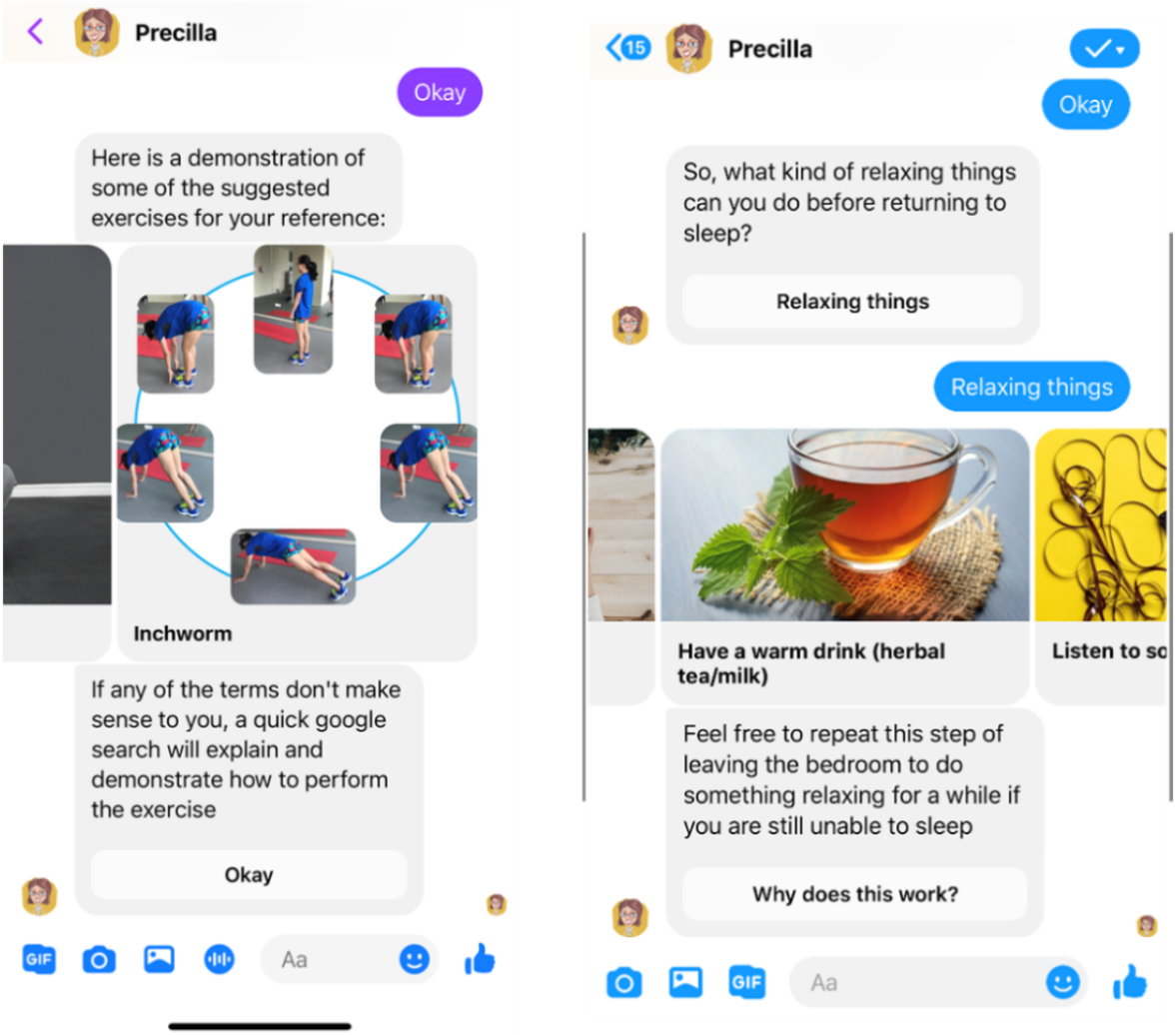
XSL•FO
RenderX

## Intervention

The conversational agent was designed to be used on Facebook Messenger using a free web-based tool, *Chatfuel* [22]. A research associate (AS) and a PhD student (DAD) developed the script for the program and performed the input. The intervention focused on diabetes and prediabetes knowledge, diet, physical activity, sleep, and stress management. These were the topics of focus identified in other diabetes prevention programs targeting lifestyle change [23-25].

The content was informed by existing evidence-based sources of information, including clinical guidelines and systematic reviews. Advice on improving sleep quality was generated from published evidence reporting on techniques and successful interventions for sleep disorders [26]. The domain on stress was informed by distance learning–based stress management techniques identified from a review of existing studies that described methods to reduce stress and improve health [27]. Pertinent nutritional advice for individuals with prediabetes was obtained from authenticated government health portals and other validated health and nutrition webpages [28]. The collated advice was then compartmentalized into themes, which translated into the topic of focus for each interaction between conversational agent and user (Multimedia Appendix 1 [8,26,27,29-53]). The content for the section on physical activity was informed by advice on the recommended duration and intensity of exercise from Singapore's Health Promotion Board [28]. Validated fitness routines were then presented as part of the conversation as methods to achieve the necessary level of fitness [29]. Examples of conversational exchanges between the conversational agent and users are presented in Figure 1.

**Figure 1.** A representation of what the interactions between Precilla and users entailed.




We also followed the Capability, Opportunity, Motivation, Behavior model of behavior to guide the development of the intervention [54]. This model posits that to exhibit a particular behavior (B), the participant must be physically and psychologically capable (C), have sufficient social and physical opportunities (O) to perform the behavior, and must have the desire or need to do so—motivation (M). This was important in determining the inclusion criteria for the intervention (eg, physically fit with no comorbidities) and in conversation designing.

The content was mostly presented in the form of text, supplemented with some images to make the conversational agent more engaging and to enhance the user experience. These images were obtained from free-to-use sources or photographs taken by our study team members. To further contribute to a positive user experience that mimicked human interaction, the



XSL•FO
RenderX

conversational agent was given a name, *Precilla*, and it displayed human-like characteristics such as in the tone of speech, profile picture, and through using the *typing* function for messages. From here onward, we use *Precilla* to refer to the conversational agent used in our pilot study.

The intervention was scheduled to last for 4 weeks, whereby participants would receive 4 messages per week, (every other day), 1 for each topic (diet, exercise, sleep, and stress). A sample of the dialog tree is presented in Multimedia Appendix 2.

## Procedure

An overview of the study workflow is demonstrated in Figure 2. Interested individuals were first required to complete an eligibility questionnaire. The eligible participants then provided informed consent, after which they were requested to complete a baseline questionnaire on a web-based tool, *REDCap* (Research Electronic Data Capture; Vanderbilt University) [55]. The questionnaire obtained general information regarding demographics as well as data on diet, physical activity, sleep, and stress.

**Figure 2.** Study workflow.

The actual pilot study lasted for 4 weeks. Upon completion of the piloting period, the participants were required to complete a follow-up questionnaire that contained all the details outlined in the baseline questionnaire, with some additional questions on conversational agent usability and their overall satisfaction with the study. Interested participants also had an opportunity to take part in a follow-up interview to share their views on the conversational agent, their experience while taking part in this web-based study, as well as their thoughts on points of improvement.

## Outcomes

### *Primary Outcomes*

The primary outcomes of interest for this study were the feasibility of recruitment and retention of participants, acceptability of the intervention, and participant engagement with the intervention.

### *Feasibility of Recruitment and Retention*

Feasibility in this study was determined by recruitment and retention. Feasibility of recruitment was defined as the ability of the researchers to recruit at least 75% (150/200) of the target sample on the web using Facebook within a 1-month period. Feasibility of retaining participants was defined as at least 32.7% (49/150) of the recruited participants completing the study.

### *Acceptability of the Intervention*

The acceptability of Precilla was measured through questions on usability and satisfaction in the follow-up questionnaire. Questions were asked on the participants' overall satisfaction with Precilla, their likelihood of using Precilla again and recommending it to others, as well as the impact of the interactions on their health.

The usability questionnaire was split into 2 sections; the questions in section 1 related to how participants perceived the content of Precilla's input. They were required to provide a rating for each statement from the following options: strongly agree, agree, neutral, disagree, strongly disagree. Questions were asked on ease of use, enjoyment, long-term use of Precilla, and language, as well as motivation to make healthier food choices, exercise, change sleeping habits, and better manage stress.

The questions in section 2 were concerned with the usability of Precilla and its associated delivery methods. The participants were provided with the same options to choose from as in section 1, and they were asked to share their opinions on the mode of communication (buttons, text, images, etc), the suitability of Facebook Messenger as a channel of communication, the number of messages sent, the timing of messages, and Precilla's personality.

### *Participants' Engagement*

A further measure of acceptability of the intervention was the participants' engagement with Precilla. Data on this aspect was collected manually by analyzing individual conversations between Precilla and the participants on Facebook Messenger. This involved noting down the duration of interactions and counting the number of complete, incomplete, and nil interactions. Immediate responses were defined as interactions made by the user within an hour of message receipt. We also collected data directly from Chatfuel analytics, such as user retention, free text typed by users (such as questions or random utterances), and the total number of attempted interactions with Precilla. In addition, the technological savviness of the study population was gauged by asking the participants to rate their own technological competency on a scale from 1 to 10.

### *Secondary Outcomes*

Our secondary outcomes were related to the efficacy of the intervention with regard to the participants' QoL, diabetes knowledge, diet, physical activity, sleep, and stress over 4 weeks.

QoL was measured using the short form version of the QoL Enjoyment and Satisfaction Questionnaire [56]. The questionnaire comprises 14 items that are rated on a 5-point scale that indicates the degree of enjoyment or satisfaction experienced during the past week. The total score for the 14 items, which cover the topics of work, social life, health, and overall well-being, ranges from 14 to 70. A percentage of the maximum score is also reported; for example, if a participant



XSL•FO
RenderX

scores 20, the percentage—29% (20/70)—is also reported. The last 2 questions on medication adherence and overall life satisfaction were reported as percentages (Multimedia Appendix 3).

The knowledge questionnaire had 3 separate sections. Section 1 was concerned with participants' knowledge about healthy living, prediabetes, and diabetes; section 2 asked for their perceptions on how healthy their lifestyle is and the likelihood of their developing diabetes; section 3 tested their knowledge of risk factors for type 2 diabetes. The questionnaire was derived from an adaptation of questions presented by the Michigan Diabetes Research Centre [57]. The questionnaire was adapted such that, in section 1, questions regarding the definitions of diabetes and prediabetes were added to test participants' knowledge of these conditions. In section 2, the wording of some questions about participants' efforts to make healthy lifestyle behavior changes in the past year was adapted to obtain more detail in their responses from just yes or no to a scaled format. In section 3, some risk factors that were not applicable to the Singaporean context were removed (eg, being Asian American, Hispanic, or African American).

The participants' diets were assessed based on an adaptation of the Food Frequency Questionnaire where questions were asked on the frequency of daily intake of vegetables, fruits, fried foods, and sweetened drinks [58]. Physical activity was assessed using the International Physical Activity Questionnaire (IPAQ), which was adapted specifically for use in this study [59]. Questions were asked on the intensity of exercise (vigorous, moderate, or light), the frequency of physical activity (in days per week), and the length of each session (in minutes). Only sessions lasting at least 10 minutes qualified as physical activity. In line with IPAQ scoring, a metabolic equivalent of task (MET) score was calculated. The MET score represented the amount of energy expended when carrying out physical activity. For consistency, walking was given a score of 3.3 METs, moderate physical activity 4 METs, and vigorous physical activity 8 METs. To calculate MET minutes per week, the MET value was multiplied by the minutes for which the activity was carried out and again by the number of days in the week that the activity was undertaken. As some participants provided a range for their responses, for consistency, a mean value was used for the calculation. For example, for number of days, that is, 1-3 days, the mean was calculated as 2. For session length, that is, 10-20 minutes, the mean was calculated as 15.

A MET score of 600 METs per week indicated that an individual was moderately physically active, whereas 1500 METs per week indicated a high level of physical activity. Any score that did not qualify as moderate or high was considered an indication of a low level of physical activity.

Sleep scoring was done using the Pittsburgh Sleep Quality Index (PSQI) questionnaire [60]. A score of 0 indicated excellent sleep quality, whereas 10 indicated severely poor sleep quality. Poor sleep quality was defined as participants with a global PSQI score higher than 5. The stress level of the participants was gauged using the Perceived Stress Scale, where it was possible to receive a score between 0 and 40 [61]. Scores ranging from 0 to 13 were considered low perceived stress, from 14 to 26 moderate perceived stress, and from 27 to 40 high perceived stress.

## Data Analysis

Data analysis was conducted on the data collected at baseline and follow-up (Table 1). The participant responses on satisfaction and usability were presented as percentages of each option on a Likert scale (eg, strongly agree, disagree, much better than before, or neither better nor worse). The participant outcomes for QoL, knowledge, physical activity, sleep, and stress were all presented as pre- and postscores using the individualized scoring system of each questionnaire.

**Table 1.** Overall summary of results (N=60).

| | Baseline values (n=60) | Follow-up (4 weeks) values (n=56) |
|---|---|---|
| **Knowledge, mean (SD; range)** | | |
| Section 1 | 7 (1.37; 6-18) | 6 (0.93; 6-18) |
| Section 2 | 28 (2.7; 12-48) | 27 (3.5; 12-48) |
| Section 3 | 10 (4.1; 8-32) | 9 (2.2; 8-32) |
| Sleep, mean (SD) | 4 (2.36) | 4 (2.45) |
| Stress, mean (SD) | 17 (5.13) | 16 (5.10) |
| Physical activity (MET[a] score), mean (SD) | 1080 (816) | 1075 (872) |
| **QoL[b]** | | |
| 14-item score, mean (SD) | 54 (6.90) | 53 (6.46) |
| Maximum score (%) | 77 | 75 |

[a]MET: metabolic equivalent of task.

[b]QoL: quality of life (measured using the short form version of the QoL Enjoyment and Satisfaction Questionnaire).

We performed descriptive analyses of the data. The data were presented using percentages, means, and SDs. As this was a feasibility study, no hypothesis testing was performed to assess the efficacy of the intervention [62]. Mean differences between baseline and follow-up were presented, accompanied by 95% CIs [62].



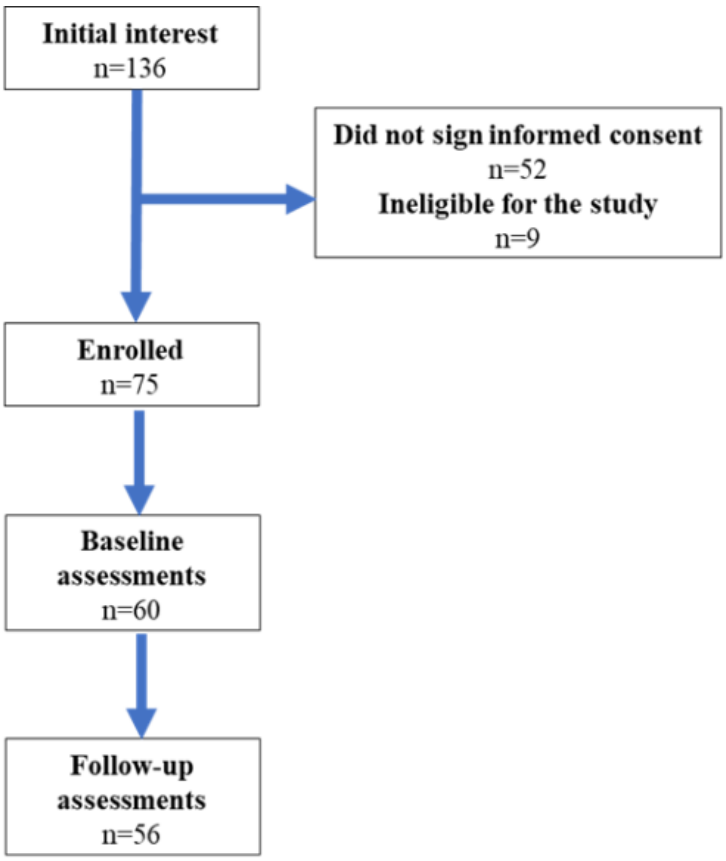
XSL•FO
RenderX

## Results

### Recruitment and Retention

A total of 136 individuals expressed initial interest in participating in the study; however, 9 (6.6%) were ineligible because of comorbidities or because they had not installed Facebook Messenger on their smartphone, and a further 52 (38.2%) completed the screening questionnaire but did not proceed to provide informed consent, leaving 75 (55.1%) participants eligible for participation. Of these 75 participants, 60 (80%) completed baseline assessments, and of these 60, 56 (93%) completed the follow-up questionnaires (Figure 3).

**Figure 3.** The flow of participants through the study.




### Characteristics of Participants

Of the enrolled participants, 38% (23/60) were men. The mean age was 33.7 years (SD 9.3), the mean BMI was 22.3 kg/m$^2$ (SD 3.8; Table 2), and the average technical competency on a scale of 1 to 10 was 8.07.



XSL•FO
RenderX

**Table 2.** Characteristics of all the enrolled participants who completed baseline assessments (N=60).

| Characteristics | Baseline values |
|---|---|
| Age (years), mean (SD) | 33.7 (9.3) |
| BMI (kg/m$^2$), mean (SD) | 22.3 (3.8) |
| Gender (male), n (%) | 23 (38) |
| **Ethnicity, n (%)** | |
| Chinese | 48 (80) |
| Malay | 5 (8) |
| Indian | 3 (5) |
| **Others** | |
| White | 2 (3) |
| Burmese | 2 (3) |
| **Marital status, n (%)** | |
| Currently married | 25 (42) |
| Never married | 33 (55) |
| Separated | 1 (2) |
| Divorced | 1 (2) |
| **Highest level of education, n (%)** | |
| University and above | 50 (83) |
| Polytechnic diploma | 1 (2) |
| Other diploma and professional qualification | 1 (2) |
| A[a]-level or NTC[b]-1 or NTC-2 or certificate in office or business skills or its equivalent | 6 (10) |
| O[a] or N[c]-level or NTC-3 certificate or its equivalent | 1 (2) |
| Secondary school | 1 (2) |
| **Work status, n (%)** | |
| Employed | 37 (62) |
| Student (full time) | 17 (28) |
| Homemaker or housewife | 3 (5) |
| Unemployed (able to work) | 2 (3) |
| Retired | 1 (2) |
| **History of parents, sibling, or child with type 2 diabetes, n (%)** | |
| Yes | 11 (18) |
| No | 49 (82) |
| **History of hypertension, n (%)** | |
| Yes | 2 (3) |
| No | 58 (97) |

[a]A-level or advanced-level examinations are taken by students at the age of 18, 2 or 3 years after completing their O-level or ordinary-level examinations, which are taken by students at the age of 16 after 4 years (or 5 years) of secondary school.

[b]NTC: National Technical Certificate.

[c]N-level: The Singapore-Cambridge General Certificate of Education Normal-level. Secondary students in Singapore can move between two streams based on their academic performance: 4 years of study culminating in the O-level (ordinary level) or the N-level (normal level) examinations. N-level students may participate in a fifth year of study to take the O-level examinations.



XSL•FO
RenderX

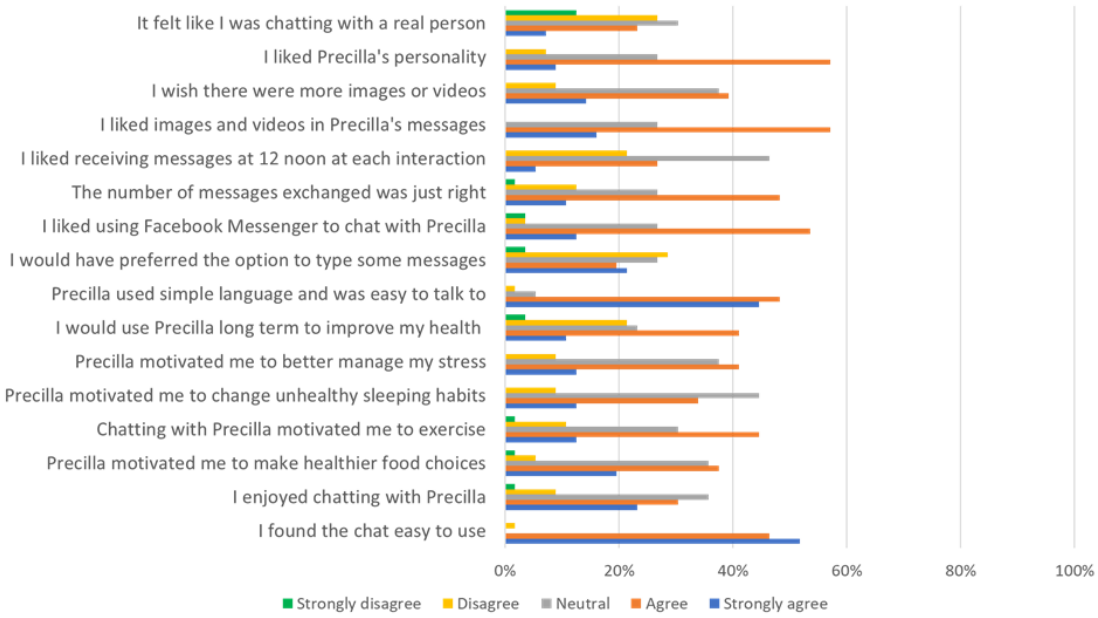
## Satisfaction

Of the 56 participants who followed the study through till completion, 52 (92%) were moderately or very satisfied with Precilla, 30 (54%) thought that their likelihood of recommending Precilla to others would be *somewhat likely*, 32 (57%) felt that the likelihood of their opting to use Precilla again for personal use would also be *somewhat likely*, and, finally, 29 (51%) posited that their health was somewhat better or much better than before.

## Usability

The *agree* and *strongly agree* responses exceeded 50% (28/56) for all the questions relating to Precilla's acceptability, except for question 5, "Chatting with Precilla motivated me to change unhealthy sleeping habits," where the collective response in agreement (agree and strongly agree) was only 46% (26/56) (Figure 4). The smallest percentage of disagreement was for questions 1 and 8, indicating that most of the participants (55/56, 98%) found the chat easy to use and thought that Precilla used simple language and was easy to talk to. The *neutral* response was a common selection by participants for questions 2-7.

**Figure 4.** Participants' opinions of Precilla's usability and acceptability.




Of the 56 participants, 37 (67%) agreed that Facebook Messenger was an appropriate medium to deliver the messages sent by Precilla. The remaining participants were either neutral or disagreed. *Telegram* and *WhatsApp* were proposed as potential alternative delivery media. Of the 56 participants, 9 (15%) thought that the number of messages exchanged with Precilla was too high, and they would have preferred the intervention to deliver fewer messages; 12 (21%) disagreed that 12 PM was the most appropriate time to receive messages, whereas the remaining 44 (79%) either agreed, strongly agreed, or were neutral (Figure 4) and indicated a preference for interacting with Precilla after work, in the evening or before bedtime. Some indicated specific timings such as 5 PM, 8 PM, or 10 PM, and 1 participant suggested 9 AM or during the morning commute to work. There were no disagreements to statement 13 ("I liked images and videos in Precilla's messages") and minimal disagreement for statements 14 (5/56, 9%) and 15 (4/56, 7%) on the inclusion of more visual components and Precilla's personality, respectively. Most participants did not find Precilla to be very human-like and could clearly tell that they were communicating with a conversational agent. Of the 56 participants, 3 (5%) suggested some degree of personalization, whereby Precilla's messages should be initiated when the user prompted the conversational agent and not the other way around.

## Conversational Agent Engagement Data

The number of complete, incomplete, and absent interactions were noted (Figure 5). In addition, the number of immediate and delayed interactions were also counted (Figure 6). Some interactions were started but not completed. The reasons for not completing interactions are presented in Multimedia Appendix 4.



XSL•FO
RenderX

**Figure 5.** Percentage of total weekly interactions that were complete, incomplete, or absent (not attempted at all).

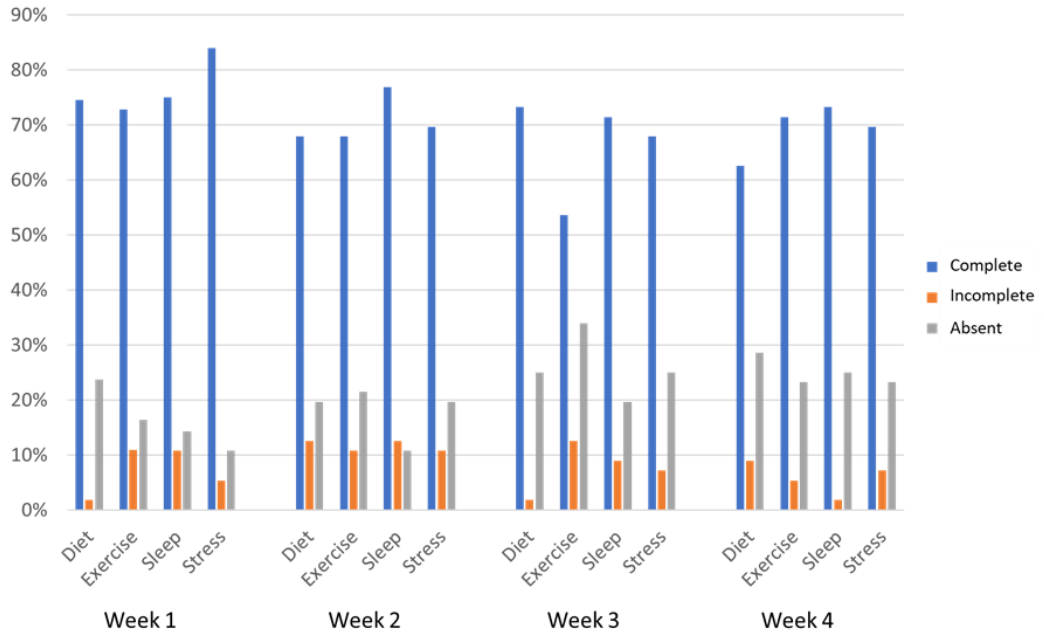
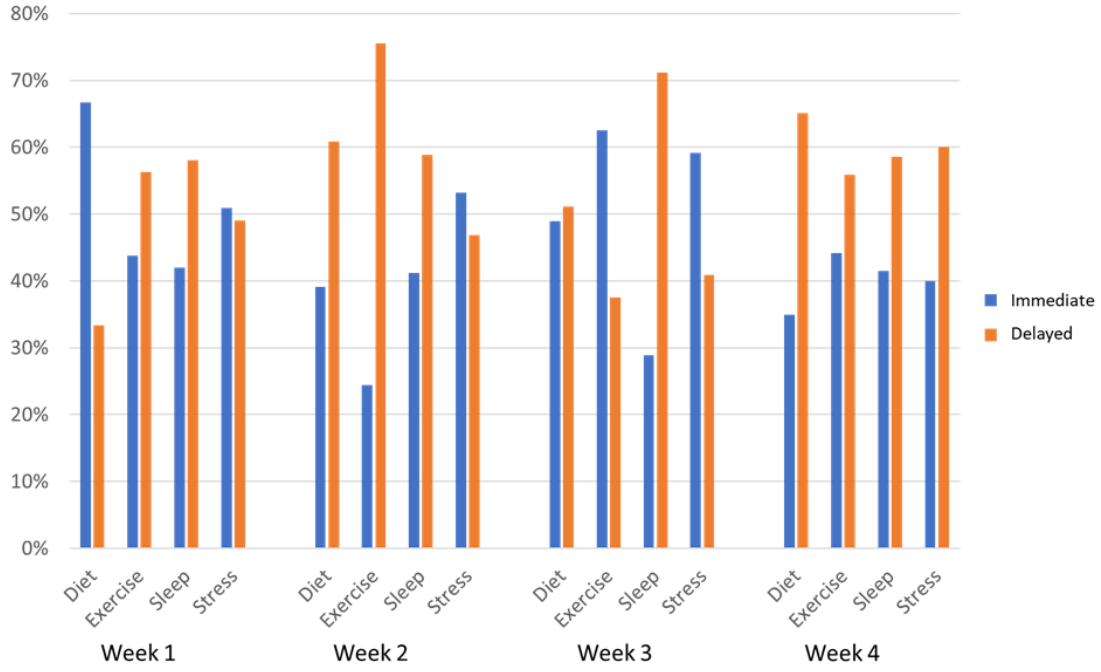



**Figure 6.** Number of immediate and delayed interactions between participants and Precilla for each category.




All the conversations were completed by at least 50% (>30/56) of the participants consistently over the period of the study. The number of participants who completed interactions from start to finish exceeded 71% (40/56) for 56% (9/16) of the interactions, denoting high engagement with the conversational agent. Similarly, at least 40% (<22/56) of the participants responded immediately to the conversational agent 75% (12/16 interactions) of the time, even in the last week of the study, denoting consistently high conversational agent engagement.

The participants were restricted to a predetermined list of options when providing responses to the questions asked or information provided. A frequent cause of disruption to the conversation flow was participants typing free text, which led them off track. This resulted in their being redirected to the main menu, which they potentially found confusing, or they were reluctant to go through the whole conversation flow again and, hence, failed to complete that interaction. The average duration of each interaction was in the 2- to 5-minute range. More data on conversational agent engagement are presented in Multimedia Appendix 4.

## QoL Score

The 14-item QoL score was 54 (SD 6.90) at baseline and 53 (SD 6.46) at follow-up (mean difference 0.95; 95% CI −1.51 to 3.41). Of the participants taking medication (10 at baseline, 17 at follow up), participants rated adherence as fair (baseline



XSL•FO
RenderX

5/10, 50%; follow-up 6/17, 35%), good (baseline 5/10, 50%; follow-up 9/17, 53%), and very good (baseline 0/10, 0%; follow-up 2/17, 12%). The participants rated overall life satisfaction as fair (baseline 16/60, 27%; follow-up 12/56, 21%), good (baseline 35/60, 58%; follow-up 33/56, 59%), or very good (baseline 9/60, 15%; follow-up 10/56, 18%), and 2% (1/56) of participants reported poor overall life satisfaction at follow-up.

### Knowledge

The scoring system for the questionnaires meant that lower scores denoted better knowledge outcomes (sections 1 and 3) and lower perceived risk of diabetes (section 2). In section 1, the scores were 7 (SD 1.37) at baseline and 6 (SD 0.93) at follow-up (mean difference 0.52, 95% CI 0.09-0.95); the possible score range was 6-18. In section 2, the scores were 28 (SD 2.7) at baseline and 27 (SD 3.5) at follow-up (mean difference 0.9, 95% CI −0.25 to 2.05); the possible score range was 12-48. In section 3, the scores were 10 (SD 4.11) at baseline and 9 (SD 2.2) at follow-up (mean difference 1.0, 95% CI −0.23 to 2.23); the possible score range was 8-32.

### Diet

The number of individuals consuming vegetables at least once a day was 27% (16/60) at baseline and 29% (16/56) at follow-up. Similarly, the percentage of participants having three portions of fruit per day was 3% (2/60) at baseline and 7% (4/56) at follow-up. With regard to intake of unhealthy food comprising sweetened beverages and fried food or snacks, the *almost never* consumption category for sweetened beverages was 38% (23/60) at baseline and 45% (25/56) at follow-up, whereas for fried food and snacks, the *almost never* consumption category was 25% (15/60) at baseline and 30% (17/56) at follow-up.

### Physical Activity

The median METs-per-week score was 857 at baseline (IQR 902) and 765 (IQR 882) at follow-up. The average categorical score remained constant from baseline to follow-up where the average exercise intensity was *moderate*. The mean time spent sitting was 439 minutes at baseline and 406 minutes at follow-up. In addition, the median amount of moderate and vigorous physical activity was the same, 30 min per week (IQR 90) at baseline and 50 min per week (IQR 90) at follow-up.

### Sleep

According to the scoring method, a lower score indicated better sleep. The mean PSQI score was 4.38 (SD 2.36) at baseline and 4.43 (SD 2.45) at follow-up (mean difference −0.05, 95% CI −0.93 to 0.83). The prevalence of poor sleep quality was 28% (16/60) at baseline and 30% (16/56) at follow-up. The participants who reported sleeping for less than 7 hours were classified as short sleepers; 82% (49/60) were short sleepers at baseline, whereas 88% (49/56) were short sleepers at follow-up.

### Stress

The scoring for stress meant that lower scores denoted lower stress levels and higher scores denoted higher stress levels. The mean score at baseline was 17 (SD 5.13), denoting a moderate stress level, which was maintained even at follow-up, mean 16 (SD 5.1; mean difference 0.73; 95% CI −1.15 to 2.61).

## Discussion

### Principal Findings

To our knowledge, this is the first web-based pilot study to test the feasibility and acceptability of a healthy lifestyle behavior change conversational agent in the Singaporean population. In this web-based study, which used Facebook Messenger as a delivery medium for a conversational agent targeting healthy lifestyle behavior change in the Singaporean population, we managed to recruit 37.5% (75/200) of the target sample size in 1 month. Retention was high at 93% (56/60); conversational agent engagement was also high, with all the conversations being completed by at least 50% (28/56) of the participants consistently. Furthermore, at least 40% (22/56) of the participants responded almost immediately 75% (12/16) of the time. Acceptability, usability, and satisfaction were also generally high. In general, we were able to conduct the study with high fidelity, and each phase ensued as planned. Any definitive signs of improvement in QoL, knowledge, diet, physical activity, sleep, and stress would have to be studied in more detail in a future study of effectiveness.

The secondary outcomes in the study were measured using validated scales, namely, the short form version of the QoL Enjoyment and Satisfaction Questionnaire, PSQI, Perceived Stress Scale, Food Frequency Questionnaire, and IPAQ. We took note that although these questionnaires have not necessarily been optimized for the Singaporean population, these well-established scales were chosen to ensure a certain level of validity and reliability in our results.

### Comparisons With Existing Literature

Conversational agents have been used for healthy behavior change; however, they have been used mainly in niche areas such as smoking cessation, alcohol misuse treatment, and the promotion of physical activity in sedentary populations [63-65]. The application presented in this study is very comprehensive, novel, and relevant to the general population in Singapore.

The outcome measures were similar to those in other studies looking at the acceptability of conversational agents in health care, and acceptability was high in other studies as well as in this feasibility trial. Previous studies noted high response rates (ie, conversational agent engagement) and strongly agree or agree scores for user-friendliness, appropriateness, consistency, and speed of response, as in our study [66]. The other measures in prior studies that denoted acceptability were high perceived ease of use, usefulness, and intention to use, similar to our measures of ease and enjoyment of use, motivating change in unhealthy habits, and intention to use Precilla again or recommend it to others, respectively [67,68].

In another study using a health behavior change conversational agent, high compliance was attributed to a rewarding game system [68]. Considering the slight decrease in weekly compliance in our feasibility study (43/56, 77% completed interactions in week 1 vs 39/56, 69% in week 4), it may be worth exploring the possibility of introducing a gaming or reward component in future iterations. Personalization has previously been met with positive user reception, for example,



XSL•FO
RenderX

by providing personalized advice to maintain target blood glucose levels or personalized reminders for taking medication [69]. Similarly, personalized content and message timing delivery should be implemented in future versions of Precilla, considering the preference indicated by users.

It has previously been indicated that users tend to prefer interacting with female conversational agents, as evidenced by Siri, Cortana, and Alexa, for example [70]. In addition, the study by Brahnam et al [71] has explained that in the field of human-computer interaction, "the standard of believability has become inextricably linked to gender personification, especially female personification." As such, we chose to use *Precilla* and reinforced her character with a profile picture.

Views on conversational agent personality have been mixed, depending on the function of the agent. For example, participants appreciated an empathetic demeanor from conversational agents for e-therapy, in contrast to the participants' indifference to Precilla's humanity in this pilot study, where she played more of an impersonal, informative or educational role [72]. It may be that the conversational agent's purpose determines how important the degree of humanity is.

Another aspect of this study was the effectiveness of participant recruitment through social media. We chose a goal of 75% (150/200) recruitment rate based on other studies involving mobile health trials having achieved an application rate of 86% (70/81) for enrollment through Facebook or an 85.1% (605/711) response rate to Facebook advertisements in 45 days [73,74]. Although our study only yielded a recruitment rate of 37.5% (75/200) in the given period, it aligns quite appropriately with the recruitment rates of 33%, 30%, or 37.7% in other web-based mobile apps or Facebook trials [75-77]. Hence, our 37.5% (75/200) recruitment rate seems to indicate feasibility regarding the recruitment of participants in the general population in Singapore for a web-based conversational agent intervention delivered through Facebook.

A systematic review accentuated the effectiveness of Facebook as a recruitment tool [78]. It was suggested that on average, studies allowed a 3-month recruitment period, and Facebook was found to be more efficient than traditional methods (print, radio, television, email, or word of mouth) because of the reduced costs, shorter recruitment times, and ability to connect with harder-to-reach populations. In future studies, we may need to consider complementing our current no-fee Facebook approach with advertisements or other recruitment methods, as well as potentially a longer recruitment period.

## Implications and Future Research

In terms of the recruitment, we noticed some attrition among the eligible participants just before provision of informed consent. As the only communication was through email from an institutional email address, the uncertainty involved in remote participation, such as never meeting the study team in person, could have been a cause for the participants' apprehension. This lack of direct engagement may have led to some participants not feeling comfortable sharing their digital signature with us because of data protection and privacy concerns. In future research studies, a mix of a digital recruitment approach and face-to-face exchange (possibly in the form of recruitment or debriefing) could provide participants with an opportunity to ask questions and validate the legitimacy of the study.

Regarding the preliminary efficacy, minimal improvements were observed in the participants' knowledge, stress, and diet, whereas there was a lack of improvement in QoL and sleep and no change in physical activity. The dearth of significant improvement in these areas could be attributed to the content and delivery not being adequately designed to target effectiveness and, possibly, the short study length (4 weeks), which may not have been sufficient for any noteworthy changes to have been observed. Furthermore, this study was not designed to test the effectiveness of the intervention, and this should be explored in more detail in future iterations.

The participants indicated that they could clearly tell that Precilla was indeed a conversational agent. This was beneficial in confirming Precilla's identity and in reducing expectations from users that the conversational agent should think, function, and react like a human being. It was interesting to note that although the participants did not find Precilla to be very human-like, they still rated the conversational agent highly in terms of content, usability, general acceptability, ease of use, visual components, and so on. This shows that, for this study population and for a conversational agent with Precilla's functions and capabilities, acceptability was not dependent on the anthropomorphism of the conversational agent.

The conversational agent engagement data revealed that there was a slight decrease in the number of interactions that the participants completed fully with each successive week. It is possible that the content and delivery methods could have come across as repetitive because we provided a review of take-home points from the previous interaction from week 2 onwards. Alternatively, perhaps the routine that Precilla followed made the intervention very regimented, not leaving much room for spontaneity in the form of varied message timings or even variations in the ways that users could respond to Precilla, for example, using free text. These factors may have contributed to the decrease in interest over time. Future iterations of Precilla could explore introducing more novelty and personalization, as recommended by the participants in the telephone interviews. The participants shared some points for improvement, including alternative modes of delivery such as WhatsApp and Telegram, a possibility of shortening the conversation lengths, and more personalized timings for message delivery.

## Strengths and Limitations

One of the main strengths of the study is the high fidelity in the delivery of this low-cost, fully web-based feasibility study. We did not experience any software malfunctions and were able to implement the intervention as per protocol. The content of the intervention was evidence based, and it was co-designed with members of the target population to make the intervention as relevant to them as possible. We managed to reach a diverse range of age groups—from 23 to 60 years—using Facebook without the need to pay for advertising. Although the recruited sample—recruited over 4 weeks using solely Facebook—consisted of fewer participants than our target size, it is comparable with those reported in other studies that used



XSL•FO
RenderX

paid Facebook advertisements within the same time frame. However, the participants were recruited through healthy living–focused Facebook groups and pages; therefore, these individuals were already very much invested in healthy lifestyle behavior change and more likely to use apps and social media. Furthermore, the participants were Facebook users, capable of navigating Facebook Messenger, had high technological competency, and a very high level of tertiary education. As such, our findings may not be generalizable to other demographic groups such as older people; individuals with lower education levels; or those not active on, or familiar with, social media. However, this study did largely cover the general characteristics of the average Singaporean, making the findings very much valuable and relevant. A further limitation is that although we used validated outcome measurement tools, all outcomes were self-reported by the participants. Furthermore, although these measurement scales are well reputed and established, they were not necessarily optimized for use in the Singaporean population. Finally, we used a single-arm pre-post study design, which is acceptable for a feasibility study but does not allow for assessment of the effectiveness of the intervention.

## Conclusions

This web-based feasibility study showed that the delivery of a conversational agent for healthy lifestyle behavior change using Facebook Messenger is, to a large extent, feasible in Singapore. Precilla is a low-cost intervention that was popular among the participants and was well received, with most participants being satisfied with the intervention and prepared to recommend it to friends and family. This study demonstrated the ability to conduct a web-based trial to assess the impact of a novel intervention. Our preliminary data on the acceptability of the intervention showed the need for further enhancement of this conversational agent intervention, potentially through humanization of the agent and personalization of the messaging. Such an intervention needs to be evaluated with a rigorous study design and larger sample size.

## Acknowledgments

This research was supported by the Ageing Research Institute for Society and Education, Nanyang Technological University, Singapore. This study was also supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise program.

## Authors' Contributions

LTC conceived the idea for this study. DAD, AS, and LTC developed the resources and administered the study. DAD recruited the participants and collected the data. DAD analyzed the data and interpreted the results with input from TS and LTC. DAD, TS, LTC, AS, JB, and YLT were involved in writing the paper and approved the submitted version.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Sources of information for the conversational agent.
[DOCX File , 19 KB - formative_v5i12e27956_app1.docx ]

Multimedia Appendix 2
Dialogue tree sample outline for introductory week. The message blocks are in grey, and answer button options in blue. User's name will be seen in the {first name} area. The "typing" lasts between 2-5 seconds depending on the length of text to be read.
[DOCX File , 86 KB - formative_v5i12e27956_app2.docx ]

Multimedia Appendix 3
Questionnaires administered to participants.
[DOCX File , 26 KB - formative_v5i12e27956_app3.docx ]

Multimedia Appendix 4
Conversational agent engagement data.
[DOCX File , 18 KB - formative_v5i12e27956_app4.docx ]

## References

1. Yates T, Davies MJ, Khunti K. Obesity and chronic disease in younger people: an unfolding crisis. Br J Gen Pract 2012 Jan;62(594):4-5 [FREE Full text] [doi: 10.3399/bjgp12X616201] [Medline: 22520657]
2. Phan TP, Alkema L, Tai ES, Tan KH, Yang Q, Lim W, et al. Forecasting the burden of type 2 diabetes in Singapore using a demographic epidemiological model of Singapore. BMJ Open Diabetes Res Care 2014 Jun 11;2(1):e000012 [FREE Full text] [doi: 10.1136/bmjdrc-2013-000012] [Medline: 25452860]



XSL•FO
RenderX

3. Lai L. One in three Singaporeans not active enough. The Straits Times. 2018. URL: https://www.straitstimes.com/singapore/one-in-three-singaporeans-not-active-enough [accessed 2021-09-20]
4. Dale AC, Nilsen TI, Vatten L, Midthjell K, Wiseth R. Diabetes mellitus and risk of fatal ischaemic heart disease by gender: 18 years follow-up of 74,914 individuals in the HUNT 1 Study. Eur Heart J 2007 Dec;28(23):2924-2929. [doi: 10.1093/eurheartj/ehm447] [Medline: 17947212]
5. Gu K, Cowie CC, Harris MI. Mortality in adults with and without diabetes in a national cohort of the U.S. population, 1971-1993. Diabetes Care 1998 Jul;21(7):1138-1145. [doi: 10.2337/diacare.21.7.1138] [Medline: 9653609]
6. Wong LY, Toh MP, Tham LW. Projection of prediabetes and diabetes population size in singapore using a dynamic markov model. J Diabetes 2017 Jan;9(1):65-75. [doi: 10.1111/1753-0407.12384] [Medline: 26849033]
7. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. The Lancet 2012 Jun 16;379(9833):2279-2290 [FREE Full text] [doi: 10.1016/S0140-6736(12)60283-9] [Medline: 22683128]
8. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, Finnish Diabetes Prevention Study Group. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 2001 May 03;344(18):1343-1350. [doi: 10.1056/NEJM200105033441801] [Medline: 11333990]
9. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 07;346(6):393-403 [FREE Full text] [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
10. Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, Vijay V, Indian Diabetes Prevention Programme (IDPP). The indian diabetes prevention programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). Diabetologia 2006 Feb;49(2):289-297. [doi: 10.1007/s00125-005-0097-z] [Medline: 16391903]
11. What is a chatbot and how to use it for your business. Start It Up. 2018. URL: https://medium.com/swlh/what-is-a-chatbot-and-how-to-use-it-for-your-business-976ec2e0a99f [accessed 2021-09-20]
12. Saeed H. Developing a chatbot? Learn the difference between AI, machine learning, and NLP. Chatbots life. 2016. URL: https://chatbotslife.com/developing-a-chatbot-learn-the-difference-between-ai-machine-learning-and-nlp-40a3f745aec4 [accessed 2021-09-20]
13. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
14. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
15. Liu S, Ko QS, Heng KQ, Ngiam KY, Feng M. Healthcare transformation in Singapore with artificial intelligence. Front Digit Health 2020 Nov 17;2:2018. [doi: 10.3389/fdgth.2020.592121]
16. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009 Nov;104(11):1792-1804. [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]
17. Khadjesari Z, Murray E, Hewitt C, Hartley S, Godfrey C. Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. Addiction 2011 Feb;106(2):267-282. [doi: 10.1111/j.1360-0443.2010.03214.x] [Medline: 21083832]
18. Bailey JV, Murray E, Rait G, Mercer CH, Morris RW, Peacock R, et al. Interactive computer-based interventions for sexual health promotion. Cochrane Database Syst Rev 2010 Sep 08(9):CD006483. [doi: 10.1002/14651858.CD006483.pub2] [Medline: 20824850]
19. Lee EE, Brecht M, Park H, Lee J, Oh KM. Web-based study for improving mammography among korean american women. J Cancer Educ 2017 Jun;32(2):257-263. [doi: 10.1007/s13187-015-0920-2] [Medline: 26423058]
20. Jennings CA, Vandelanotte C, Caperchione CM, Mummery WK. Effectiveness of a web-based physical activity intervention for adults with type 2 diabetes-a randomised controlled trial. Prev Med 2014 Mar;60:33-40. [doi: 10.1016/j.ypmed.2013.12.011] [Medline: 24345601]
21. Harris J, Felix L, Miners A, Murray E, Michie S, Ferguson E, et al. Adaptive e-learning to improve dietary behaviour: a systematic review and cost-effectiveness analysis. Health Technol Assess 2011 Oct;15(37):1-160 [FREE Full text] [doi: 10.3310/hta15370] [Medline: 22030014]
22. Chatfuel Homepage. Chatfuel. URL: https://chatfuel.com/ [accessed 2021-09-20]
23. Lifestyle change program details. National diabetes prevention program CDC. URL: https://www.cdc.gov/diabetes/prevention/lcp-details.html [accessed 2021-09-20]
24. Improving lives through digital therapeutics. Cappa. URL: https://cappahealth.com/ [accessed 2021-09-20]
25. Lindström J, Louheranta A, Mannelin M, Rastas M, Salminen V, Eriksson J, Finnish Diabetes Prevention Study Group. The Finnish Diabetes Prevention Study (DPS): lifestyle intervention and 3-year results on diet and physical activity. Diabetes Care 2003 Dec;26(12):3230-3236. [doi: 10.2337/diacare.26.12.3230] [Medline: 14633807]




XSL•FO
RenderX

26. Freeman D, Sheaves B, Goodwin G, Yu L, Nickless A, Harrison P, et al. The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. Lancet Psychiatry 2017 Oct;4(10):749-758 [FREE Full text] [doi: 10.1016/S2215-0366(17)30328-0] [Medline: 28888927]
27. Blázquez Martín D, De La Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and controlling stress using mhealth: systematic search in app stores. JMIR Mhealth Uhealth 2018 May 09;6(5):e111 [FREE Full text] [doi: 10.2196/mhealth.8866] [Medline: 29743152]
28. Aim for 150 minutes of physical activity every week. Health hub. URL: https://www.healthhub.sg/live-healthy/386/150minutes [accessed 2021-09-20]
29. Heinrich KM, Patel PM, O'Neal JL, Heinrich BS. High-intensity compared to moderate-intensity training for exercise initiation, enjoyment, adherence, and intentions: an intervention study. BMC Public Health 2014 Aug 03;14(1):789 [FREE Full text] [doi: 10.1186/1471-2458-14-789] [Medline: 25086646]
30. War on diabetes: Singapore to host global conference on tackling epidemic. Channel News Asia. 2018. URL: https://www.channelnewsasia.com/singapore/diabetes-singapore-to-host-global-ministerial-conference-922466 [accessed 2021-10-04]
31. What is prediabetes? Joslin Diabetes Center, Inc. 2019. URL: https://www.joslin.org/patient-care/diabetes-education/diabetes-learning-center/what-pre-diabetes# [accessed 2021-10-04]
32. Prediabetes. Mayo Clinic. 2020. URL: https://www.mayoclinic.org/diseases-conditions/prediabetes/diagnosis-treatment/drc-20355284 [accessed 2021-10-04]
33. George AM, Jacob AG, Fogelfeld LW. Lean diabetes mellitus: An emerging entity in the era of obesity. World J Diabetes 2015 May 15;6(4):613-620 [FREE Full text] [doi: 10.4239/wjd.v6.i4.613] [Medline: 25987958]
34. Pan A, Sun Q, Manson JE, Willett WC, Hu FB. Walnut consumption is associated with lower risk of type 2 diabetes in women. J Nutr 2013 Apr;143(4):512-518 [FREE Full text] [doi: 10.3945/jn.112.172171] [Medline: 23427333]
35. Kalergis M, Yinko SL, Nedelcu R. Dairy products and prevention of type 2 diabetes: implications for research and practice. Front Endocrinol (Lausanne) 2013;4:90 [FREE Full text] [doi: 10.3389/fendo.2013.00090] [Medline: 23888154]
36. Nguyen V, Cooper L, Lowndes J, Melanson K, Angelopoulos TJ, Rippe JM, et al. Popcorn is more satiating than potato chips in normal-weight adults. Nutr J 2012 Sep 14;11:71 [FREE Full text] [doi: 10.1186/1475-2891-11-71] [Medline: 22978828]
37. Carter P, Gray LJ, Troughton J, Khunti K, Davies MJ. Fruit and vegetable intake and incidence of type 2 diabetes mellitus: systematic review and meta-analysis. Br Med J 2010 Aug 18;341:c4229 [FREE Full text] [doi: 10.1136/bmj.c4229] [Medline: 20724400]
38. Bazzano LA, Li TY, Joshipura KJ, Hu FB. Intake of fruit, vegetables, and fruit juices and risk of diabetes in women. Diabetes Care 2008 Jul;31(7):1311-1317 [FREE Full text] [doi: 10.2337/dc08-0080] [Medline: 18390796]
39. Salas-Salvadó J, Martinez-González M, Bulló M, Ros E. The role of diet in the prevention of type 2 diabetes. Nutr Metab Cardiovasc Dis 2011 Sep;21 Suppl 2:32-48. [doi: 10.1016/j.numecd.2011.03.009] [Medline: 21745730]
40. Katsanis J, Newman-Smith KC. Sleepless no more: techniques and interventions for sleep disorders. Athens J Health 2015 Feb 28;2(1):9-20. [doi: 10.30958/ajh.2-1-1]
41. Feng F, Zhang Y, Hou J, Cai J, Jiang Q, Li X, et al. Can music improve sleep quality in adults with primary insomnia? A systematic review and network meta-analysis. Int J Nurs Stud 2018 Jan;77:189-196. [doi: 10.1016/j.ijnurstu.2017.10.011] [Medline: 29100201]
42. Insomnia: relaxation techniques and sleeping habits. Informed Health. 2017. URL: https://www.ncbi.nlm.nih.gov/books/NBK279320/ [accessed 2021-10-04]
43. Too early to get up, too late to get back to sleep. Harvard Health Publishing. 2020. URL: https://www.health.harvard.edu/staying-healthy/too-early-to-get-up-too-late-to-get-back-to-sleep [accessed 2021-10-04]
44. Miller CB, Espie CA, Epstein DR, Friedman L, Morin CM, Pigeon WR, et al. The evidence base of sleep restriction therapy for treating insomnia disorder. Sleep Med Rev 2014 Oct;18(5):415-424. [doi: 10.1016/j.smrv.2014.01.006] [Medline: 24629826]
45. Colberg SR, Sigal RJ, Yardley JE, Riddell MC, Dunstan DW, Dempsey PC, et al. Physical activity/exercise and diabetes: a position statement of the american diabetes association. Diabetes Care 2016 Nov;39(11):2065-2079 [FREE Full text] [doi: 10.2337/dc16-1728] [Medline: 27926890]
46. Exercise for diabetes and get a leg up. American Diabetes Association. 2020. URL: https://www.diabetes.org/fitness [accessed 2021-10-04]
47. Benefits of exercise and physical activity. Allina Health. 2019. URL: https://www.allinahealth.org/health-conditions-and-treatments/health-library/patient-education/prediabetes/exercise-and-physical-activity [accessed 2021-10-04]
48. Migala J. Reversing diabetes through weight loss: how much should you lose? Everyday Health. 2017. URL: https://www.everydayhealth.com/type-2-diabetes/treatment/can-type-2-diabetes-reversed-through-weight-loss/ [accessed 2021-10-04]
49. Different types of physical activity. Health Hub, Ministry of Health, Singapore. 2020. URL: https://www.healthhub.sg/live-healthy/387/TypesOfActivities [accessed 2021-10-04]




XSL•FO
RenderX

50. Christmann CA, Hoffmann A, Bleser G. Stress management apps with regard to emotion-focused coping and behavior change techniques: a content analysis. JMIR Mhealth Uhealth 2017 Feb 23;5(2):e22 [FREE Full text] [doi: 10.2196/mhealth.6471] [Medline: 28232299]
51. Stress management: breathing exercises for relaxation. Michigan Medicine, University of Michigan Health. 2020. URL: https://www.uofmhealth.org/health-library/uz2255 [accessed 2021-10-04]
52. Getting started with mindfulness. Mindful. 2020. URL: https://www.mindful.org/meditation/mindfulness-getting-started/ [accessed 2021-10-04]
53. Varvogli L, Darviri C. Stress Management Techniques: evidence-based procedures that reduce stress and promote health. Health Sci J 2011;5(2):74-89 [FREE Full text]
54. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
55. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
56. Stevanovic D. Quality of Life Enjoyment and Satisfaction Questionnaire-short form for quality of life assessments in clinical practice: a psychometric study. J Psychiatr Ment Health Nurs 2011 Oct;18(8):744-750. [doi: 10.1111/j.1365-2850.2011.01735.x] [Medline: 21896118]
57. The next set of questions ask about what you think of your risk or chance for getting diabetes. Michigan Diabetes Research Centre. URL: http://www.med.umich.edu/borc/profs/documents/svi/RPS-DDwritten.pdf [accessed 2021-09-20]
58. Food frequency questionnaires. NIHR Cambridge Biomedical Research Centre. URL: https://dapa-toolkit.mrc.ac.uk/diet/subjective-methods/food-frequency-questionnaire [accessed 2021-09-20]
59. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (ipaq). e International Physical Activity Questionnaire. 2004. URL: https://www.physio-pedia.com/images/c/c7/Quidelines_for_interpreting_the_IPAQ.pdf [accessed 2021-09-20]
60. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
61. Cohen S, Kamarck T, Mermelstein R. Perceived stress scale. APA PsycTests 1983;10(2):5. [doi: 10.1037/t02889-000]
62. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
63. Elmasri D, Maeder A. A Conversational Agent for an Online Mental Health Intervention. In: Brain Informatics and Health. Cham: Springer; 2016.
64. Wang H, Zhang Q, Ip M, Fai Lau JT. Social media–based conversational agents for health management and interventions. Computer 2018 Aug;51(8):26-33. [doi: 10.1109/mc.2018.3191249]
65. Kocielnik R, Xiao L, Avrahami D, Hsieh G. Reflection companion: a conversational system for engaging users in reflection on physical activity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Jun;2(2):1-26. [doi: 10.1145/3214273]
66. Comendador BE, Francisco BM, Medenilla JS, Nacion SM, Serac TB. Pharmabot: a pediatric generic medicine consultant chatbot. J Autom Control Eng 2015;3(2):137-140. [doi: 10.12720/joace.3.2.137-140]
67. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform MobileCoach. In: Designing the Digital Transformation. Cham: Springer; 2017.
68. L'Allemand D, Shih C, Heldt K, Buchter D, Brogle B, Ruegger D. Design and interim evaluation of a smartphone app for overweight adolescents using a behavioural health intervention platform. Obesity Reviews (Suppl 1) 2018:102.
69. Cheng A, Raghavaraju V, Kanugo J, Handrianto Y, Shang Y. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. In: Proceedings of the 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC). 2018 Presented at: 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC); Jan 12-15, 2018; Las Vegas, NV, USA. [doi: 10.1109/ccnc.2018.8319283]
70. Saarem A. Why would I talk to you. Norwegian University of Science and Technology. URL: https://www.ntnu.edu/documents/139799/1273574286/TPD4505.AnneCathrine.Saarem.pdf/c440276a-8c77-41f0-8f97-dc8e47a34a9d [accessed 2021-07-17]
71. Brahnam S, Weaver M. Re/framing virtual conversational partners: a feminist critique and tentative move towards a new design paradigm. In: Design, User Experience, and Usability: Users and Interactions. Cham: Springer; 2015.
72. Griol D, Molina J, Callejas Z. Towards emotionally sensitive conversational interfaces for e-therapy. In: Artificial Computation in Biology and Medicine. Cham: Springer; 2015.
73. Teo AR, Liebow SB, Chan B, Dobscha SK, Graham AL. Reaching those at risk for psychiatric disorders and suicidal ideation: facebook advertisements to recruit military veterans. JMIR Ment Health 2018 Jul 05;5(3):e10078 [FREE Full text] [doi: 10.2196/10078] [Medline: 29980498]




XSL•FO
RenderX

74. Kayrouz R, Dear BF, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. Internet Interv 2016 May;4:1-10 [FREE Full text] [doi: 10.1016/j.invent.2016.01.001] [Medline: 30135786]
75. Kwan RY, Lee D, Lee PH, Tse M, Cheung DS, Thiamwong L, et al. Effects of an mhealth brisk walking intervention on increasing physical activity in older people with cognitive frailty: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e16596 [FREE Full text] [doi: 10.2196/16596] [Medline: 32735218]
76. Sharman A, Zhussupov B, Sharman D, Kim I. Evaluating mobile apps and biosensing devices to monitor physical activity and respiratory function in smokers with and without respiratory symptoms or chronic obstructive pulmonary disease: protocol for a proof-of-concept, open-label, feasibility study. JMIR Res Protoc 2020 Mar 26;9(3):e16461 [FREE Full text] [doi: 10.2196/16461] [Medline: 32213479]
77. Carter-Harris L, Bartlett Ellis R, Warrick A, Rawl S. Beyond traditional newspaper advertisement: leveraging facebook-targeted advertisement to recruit long-term smokers for research. J Med Internet Res 2016 Jun 15;18(6):e117 [FREE Full text] [doi: 10.2196/jmir.5502] [Medline: 27306780]
78. Whitaker C, Stevelink S, Fear N. The use of facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]

## Abbreviations

**IPAQ:** International Physical Activity Questionnaire
**MET:** metabolic equivalent of task
**PSQI:** Pittsburgh Sleep Quality Index
**QoL:** quality of life
**REDCap:** Research Electronic Data Capture

*Edited by G Eysenbach; submitted 15.02.21; peer-reviewed by W Peng, F Ghezelbash; comments to author 30.07.21; revised version received 09.08.21; accepted 24.08.21; published 03.12.21.*

*Please cite as:*
*Dhinagaran DA, Sathish T, Soong A, Theng YL, Best J, Tudor Car L*
*Conversational Agent for Healthy Lifestyle Behavior Change: Web-Based Feasibility Study*
*JMIR Form Res 2021;5(12):e27956*
*URL: https://formative.jmir.org/2021/12/e27956*
*doi:10.2196/27956*
*PMID:34870611*

©Dhakshenya Ardhithy Dhinagaran, Thirunavukkarasu Sathish, AiJia Soong, Yin-Leng Theng, James Best, Lorainne Tudor Car. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Patient-Reported Outcome Measure for Real-time Symptom Assessment in Women With Endometriosis: Focus Group Study

Esther van Barneveld[1], MD; Arianne Lim[1,2], MD, PhD; Nehalennia van Hanegem[3], MD, PhD; Lisa Vork[4,5], MD; Alexandra Herrewegh[6,7], MD; Mikal van Poll[8], MD; Jessica Manders[8], MD; Frits van Osch[5,9], PhD; Wilbert Spaans[1,2], MD; Gommert van Koeveringe[6,7], MD, PhD; Desiree Vrijens[6,7], MD, PhD; Joanna Kruimel[4,5], MD, PhD; Marlies Bongers[1,2,10], MD, PhD; Carsten Leue[7,11], MD, PhD

[1]Department of Gynaecology and Obstetrics, Maastricht University Medical Centre+, Maastricht, Netherlands
[2]Grow School of Oncology and Developmental Biology, Maastricht University, Maastricht, Netherlands
[3]Department of Gynaecology and Obstetrics, University Medical Centre Utrecht, Utrecht, Netherlands
[4]Department of Internal Medicine, Division of Gastroenterology and Hepatology, Maastricht University Medical Centre+, Maastricht, Netherlands
[5]NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, Netherlands
[6]Department of Urology, Maastricht University Medical Centre+, Maastricht, Netherlands
[7]School for Mental Health and Neuroscience (MeHNS), Maastricht University, Maastricht, Netherlands
[8]Maastricht University, Maastricht, Netherlands
[9]Department of Clinical Epidemiology, VieCuri Medical Centre, Venlo, Netherlands
[10]Department of Gynaecology and Obstetrics, Máxima Medical Centre, Veldhoven, Netherlands
[11]Department of Psychiatry and Psychology, Maastricht University Medical Centre+, Maastricht, Netherlands

**Corresponding Author:**
Esther van Barneveld, MD
Department of Gynaecology and Obstetrics
Maastricht University Medical Centre+
P. Debyelaan 25
Maastricht, 6229 HX
Netherlands
Phone: 31 433876543
Email: esthervb@gmail.com

## Abstract

**Background:** Symptoms related to endometriosis have a significant impact on the quality of life, and symptoms often recur. The experience sampling method (ESM), a digital questioning method characterized by randomly repeated momentary assessments, has several advantages over traditionally used measurements, including the ability to assess the temporal relationship between variables such as physical, mental, and social factors.

**Objective:** The aim of this study is to develop an ESM tool for patients with endometriosis to accurately measure symptoms and their course over time, allowing for personalized treatment and adequate monitoring of treatment efficacy in individual patients.

**Methods:** On the basis of international guidelines, items from validated questionnaires were selected through a literature review and during focus groups and multidisciplinary expert meetings. Data analysis was conducted using ATLAS.ti (ATLAS.ti Scientific Software Development GmbH). The feasibility and usability of the newly developed momentary assessment tool were tested for 28 consecutive days in 5 patients with endometriosis-related pain symptoms.

**Results:** Momentary assessment items contained questions concerning endometriosis symptoms, general somatic symptoms, psychological symptoms, contextual information, and the use of food and medication. A morning questionnaire on sleep and sexuality was included. In a pilot study, the patients considered the tool easy to use but time consuming. The average compliance rate of momentary assessments was 37.8% (106/280), with the highest completion rate during the first week (39/70, 56%). Therefore, it is advisable to use the ESM for a maximum of 7 days.



XSL•FO
RenderX

**Conclusions:** A new digital tool for endometriosis symptom assessment was developed using the ESM, which may help overcome the limitations of current retrospective questionnaires. After validation and testing, future studies will be planned to evaluate the use of this tool in a clinical setting in order to propose a personalized treatment plan for women with endometriosis.

***(JMIR Form Res 2021;5(12):e28782)*** doi:10.2196/28782

**KEYWORDS**

endometriosis; pelvic pain; positive affect; negative affect; patient-reported outcome measure; focus groups; experience sampling method; momentary symptom assessment; mobile phone

# Introduction

## Background

Endometriosis is defined as an estrogen-dependent condition involving the endometrium-like tissue outside the uterus [1]. It is estimated to be prevalent among approximately 10% in women of reproductive age and up to 50% in women with chronic pelvic pain (CPP) or fertility problems [2,3]. Dysmenorrhea, CPP, dyspareunia, fatigue, and infertility are the leading symptoms [4,5], which have a significant social and psychological impact, decreasing the quality of life of the patients [6-8]. Furthermore, the annual economic burden of women with endometriosis in European countries is high and similar to that of other chronic conditions [8]. The severity of the disease, as well as pelvic pain, infertility, and a higher number of years since diagnosis, are associated with higher costs of societal relevance given that these symptoms affect physical, mental, sexual, and social well-being, as well as work productivity [8-10].

Endometriosis is currently managed by surgical or medical interventions; however, approximately 50% of women have recurrent symptoms over a period of 5 years [11]. Moreover, the extent of endometriosis is not directly related to the degree of the symptoms [12], which suggests that the perception of symptoms may also be influenced by psychological and emotional distress [13-15].

## Objective

Stratified and more individualized therapeutic approaches are needed to maximize treatment efficacy and improve physical, mental, sexual, and social well-being [8-10,16]. To do so, a reliable assessment of endometriosis-related symptoms is essential. Current guidelines for symptom assessment in patients with endometriosis include the recommendations of the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials [17] and the American Society for Reproductive Medicine [18]. The former has made recommendations for clinical outcomes in pain trials [17], including pain measured in 0 to 10 scales, physical functioning, emotional functioning, symptoms, and adverse events. For patients with endometriosis, the American Society for Reproductive Medicine [18] recommends daily ratings of pelvic pain, daily ratings of dysmenorrhea, and the Endometriosis Health Profile-30 (EHP-30) [19]. Currently, there is no available assessment tool for all contextual factors that could influence endometriosis complaints, including symptom triggers and overlapping symptoms with other comorbidities. Furthermore, validated questionnaires such as the EHP-30 are retrospective. The experience sampling method (ESM) is an electronic questioning method characterized by randomly repeated self-reports on symptoms, activities, emotions, or other elements of real-time daily life [20]. This momentary assessment method has several advantages, including the ability to assess the temporal relationship between variables, high ecological validity, and highly detailed information on the experiences of the subjects. This method aims to provide self-insight, personalized treatment approaches, and adequate monitoring of the effectiveness of these treatments in individual patients. Usually, this method is made available by the use of a mobile app [20-22].

Following the previous development of an ESM tool for psychiatric conditions [20], irritable bowel syndrome (IBS) [22-24], functional dyspepsia [25], and overactive bladder syndrome [26], we aimed to develop an ESM assessment tool for patients with endometriosis.

# Methods

## Overview

This study was conducted between August 2018 and September 2019 and consisted of 5 consecutive phases: initial item selection, focus group interviews to consider input from the patients, critical evaluation through expert meetings, development of the smartphone app, and a pilot study to evaluate feasibility and usability. This study was approved by the Institutional Review Board Ethics Committee of the Maastricht University Medical Centre (MUMC+), Maastricht, the Netherlands (Ref 2018-0674; 2019-1069), and the Máxima Medical Centre, Veldhoven, the Netherlands (Ref 18.122; L19.048).

## Phase I: Question Selection

In agreement with the guidelines of the Food and Drug Administration on patient-reported outcome measure (PROM) development, item selection for the questionnaire started with an initial draft on the basis of the literature of validated outcome measures [27,28]. ESM-specific items concerning psychological, social, and environmental factors were derived from previous ESM validation studies [20-23]. Disease-specific items concerning the quality of life, affective symptoms, and disease-specific symptoms were derived from validated retrospective questionnaires (the Short Form-36, EHP-30, European Quality of Life-5 Dimensions, Generalized Anxiety Disorder-7, Patient Health Questionnaire-9, and Gastrointestinal Symptom Rating Scale-IBS). A list was created with all potentially relevant items from these questionnaires. The phrasing of the items was adjusted to conform to the momentary aspects of ESM assessments. The complete list of items was



XSL•FO
RenderX

discussed with a multidisciplinary expert team consisting of gynecologists, endometriosis experts, urologists, a psychiatrist, a gastroenterologist, and a representative of the Dutch endometriosis patient organization. All the items were discussed for potential relevance. In addition, the experts were asked in an open discussion whether there were any relevant items missing according to their field of expertise.

## Phase II: Focus Groups

### *Focus Group Recruitment*

Premenopausal patients with endometriosis (diagnosed by physical examination and imaging techniques or laparoscopy) aged ≥18 years were recruited by gynecologists from the ward of the outpatient gynecology department at the MUMC+ or the Máxima Medical Center. Furthermore, patients were recruited through advertisements on the Dutch endometriosis foundation website. Pregnant women and patients with any organic explanation for CPP besides endometriosis were not eligible for participation. Furthermore, participants had to be able to speak and understand written Dutch, as the focus groups were conducted in this language. Written informed consent was obtained from all participants before the study.

### *Focus Group Organization*

The focus groups were conducted according to the international PROM development guidelines [27] and the literature on focus group interviews [28]. For each focus group, 6 to 10 patients were invited, and 90-minute sessions were scheduled. The focus groups were conducted in 2 meeting phases according to the focus group guidelines [28], with the guidance of a moderator (EB) and at least one assistant moderator (AL, MP). In the first meeting phase, an open discussion, the participants were instructed to bring forward every item they considered essential for use in a real-time symptom assessment tool. In the second meeting phase, all items derived from the initial draft instrument were discussed in a structured manner. The patients could confirm or criticize the item value for momentary assessments and discussed the phrasing of the questions and the answer options. The focus groups were scheduled one meeting by one until saturation of input was reached, that is, the moment that the meetings no longer contributed any new items or information [25,26].

### *Statistical Analysis*

The focus group discussions were voice-recorded and transcribed (JM). Data were qualitatively reviewed and systematically analyzed using ATLAS.ti software (ATLAS.ti Scientific Software Development GmbH; workbench for the qualitative analysis of large bodies of data, eg, textual, audio, and video). Each item was grouped by domain, and all domain items were clustered. When synonyms of items were used, the most frequently mentioned item was selected for the questionnaire.

## Phase III: Expert Meeting

A final meeting with a multidisciplinary expert team (*Phase I: Question Selection*) was arranged to select the items to be used in the final questionnaire. The primary goal of the expert meeting was to critically discuss and convert the findings from the focus groups to generate applicable questions for clinical practice. A second goal was to shorten the list of ESM items to minimize response fatigue and, therefore, noncompliance of patients. All items that were included after the ATLAS.ti analysis of the focus group data were discussed for relevance until a majority was reached. In addition, the experts were asked in an open discussion whether there were any relevant items missing according to their field of expertise (ie, urology, gastroenterology, psychiatry, and gynecology).

## Phase IV: Development of a Smartphone App

The smartphone app MEASuRE (Maastricht Electronic Abdominal Symptom Reporting) was previously created by MEMIC, the center for data and information management at the Faculty of Health, Medicine, and Life Sciences of the Maastricht University and the MUMC+. The app can measure real-time experiences in daily life using the concept of ESM. MEASuRE has been described in previous research and has been adjusted for patients with endometriosis using the questions that were selected in the final expert meeting [22-26].

## Phase V: Pilot Study

The usability of the MEASuRE app has been thoroughly tested in patients with IBS. However, as we adapted the questions to an endometriosis-specific tool, we decided to conduct a pilot study with 5 patients with endometriosis to test the feasibility and usability of these changes to the tool. Given that endometriosis symptoms fluctuate during the menstrual cycle, we aimed to test whether collecting ESM data for 28 consecutive days was feasible [4,10]. Premenopausal women aged at least 18 years and diagnosed with endometriosis were recruited via the ward of the outpatient gynecology department of the MUMC+ or the Máxima Medical Center. The inclusion and exclusion criteria were similar to those in phase II, and written informed consent was obtained before participation. During the study period, ESM assessments were conducted on the patients' smartphones using the MEASuRE app. Because the sampling procedure should cover a range of waking hours and activities, the momentary assessments started after 7:30 AM and finished before 10:30 PM. The app sent out a notification at 10 random moments during the day, each within a 90-minute time frame, after which the patients could complete the identical electronic self-reports. To minimize the extent to which data were influenced by retrospective biases, the participants had to respond to the notification within the requested time frame (10 minutes). After this period, which has also been described in other studies [20,29], it was no longer possible to start the assessment. Past research has typically used 5 to 10 assessments per day to measure real-time experiences in daily life [29,30]. As missing entries were expected, we also analyzed the rates of compliance of at least 3 out of 10 assessments each day. The participants were called on the second study day to check for technical difficulties and to ensure that the questions were clear. The patients were called and interviewed after 2 weeks and at the end of the pilot study to collect feedback concerning the logistics, usability, and content of the questionnaire.



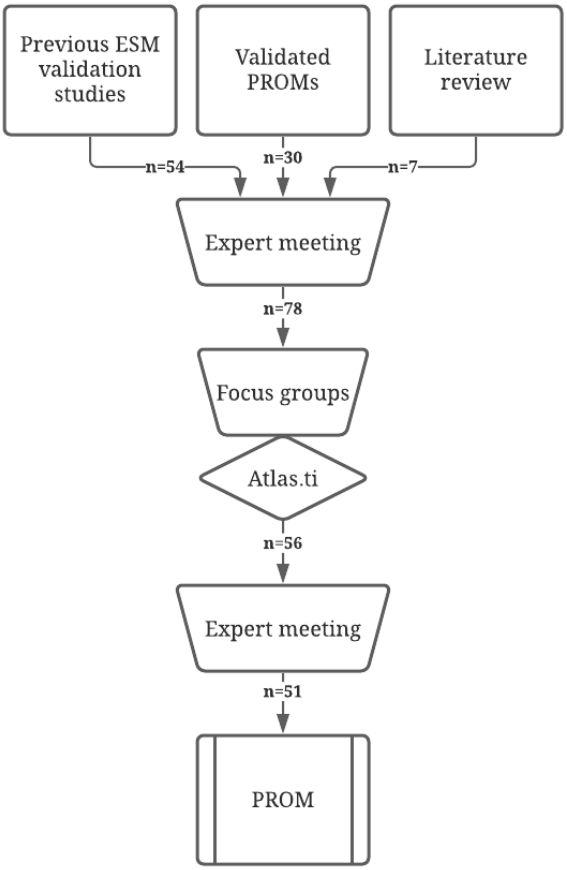
XSL•FO
RenderX

## Results

### Phase I: Question Selection

Figure 1 systematically describes the development of a momentary PROM. During question selection, 54 items concerning psychological, social, and environmental factors were derived from questions used in previous ESM validation studies [20-23], whereas 30 items were derived from validated retrospective questionnaires (the Short Form-36, EHP-30, European Quality of Life-5 Dimensions, Generalized Anxiety Disorder-7, Patient Health Questionnaire-9, and Gastrointestinal Symptom Rating Scale-IBS) and made suitable for momentary assessment. Seven questions regarding physical and endometriosis-specific symptoms were added through a clinical literature search [1-5]. During the expert meeting, 13 items were excluded on the basis of relevance. Validated scales such as the Bristol Stool Chart (used in the ESM tool for patients with IBS) and a urological urgency scale were added to make it possible to compare data from patients with endometriosis and patients with other chronic abdominal pain [31,32]. A total of 78 ESM questions were selected concerning different domains: endometriosis-specific symptoms, general somatic symptoms, sleep, sexuality, mood and psychological factors, social and contextual factors, and use of nutrition and medication.

**Figure 1.** Process of patient-reported outcome measure development. ESM: experienced sampling method; PROM: patient reported outcome measure.




### Phase II: Focus Groups

#### *Overview*

The characteristics of the women who participated in the focus group meetings are summarized in Table 1. A total of 19 patients initially agreed to participate in the focus groups; however, only 14 were present. The reasons for cancelation were illness (n=2), other plans (n=1), and family-related issues (n=1). One patient did not report any reason for the cancelation. The age of the participants ranged from 23 to 41 years. Saturation of input was reached after 3 focus groups. After the ATLAS.ti analysis of focus group data, the questionnaire comprised 56 items.



XSL•FO
RenderX

**Table 1.** Baseline characteristics.

| Variable | Study group | |
|---|---|---|
| | Focus groups (n=14) | Pilot study (n=5) |
| **Sociodemographic** | | |
| Age (years), mean (SD) | 37.1 (6.7) | 35.6 (5.6) |
| **Level of education, n (%)** | | |
| High school | 1 (7) | 0 (0) |
| College or university | 13 (93) | 5 (100) |
| **Occupational status, n (%)** | | |
| Student | 1 (7) | 0 (0) |
| Unemployed | 3 (21) | 2 (40) |
| Employed | 10 (72) | 3 (60) |
| **Relationship status, n (%)** | | |
| Single | 1 (7) | 1 (20) |
| In relationship | 13 (93) | 4 (80) |
| **Anthropometric** | | |
| BMI (kg/m$^2$), mean (SD) | 27.6 (4.7) | 26.5 (5.7) |
| **Medical, n (%)** | | |
| **Use of hormonal medication** | 12 (86) | 4 (80) |
| Oral contraceptives | 4 (29) | 2 (40) |
| Mirena IUD[a] | 4 (29) | 0 (0) |
| Progestins | 1 (7) | 0 (0) |
| GnRH[b] | 4 (29) | 2 (40) |
| Regular use of pain medication | 11 (79) | 5 (100) |
| Surgery for endometriosis | 11 (79) | 3 (60) |
| Infertility | 5 (36) | 1 (20) |
| Use of psychiatric medication | 2 (14) | 0 (0) |
| Traumatic life event in past | 3 (21) | 1 (20) |

[a]IUD: intrauterine device.

[b]GnRH: gonadotrophin-releasing hormone.

### *Morning Questionnaire*

Women with endometriosis and deep dyspareunia have been found to have lower sexual quality of life, presenting with impaired sexual functioning and decreased satisfaction, which, in turn, can negatively affect personal relationships [33]. Questions concerning sexual activity or avoidance were adapted from the modular dimension *Sexual intercourse* of the EHP-30. Furthermore, the patients considered questions regarding sleep relevant to the general state of well-being. These questions were added to the morning questionnaire, as it was considered unnecessary to assess these items repeatedly during the day [34].

### *Momentary Assessments*

Most of the answer options were presented in the numeric rating scale from 0 to 10. However, some questions had answer options on a scale of −5 to +5. The list of questions was shortened by creating subquestions in the case of positive answers. In this matter, questions regarding sexual intercourse, urination, and defecation were asked retrospectively to check whether or not they occurred. If complaints arose, the follow-up questions were asked. The patients stated that the extent of vaginal blood loss was an important issue; however, they also noted that, in the case of absence of a menstrual cycle or after hysterectomy, they did not like to answer any questions regarding blood loss. This was solved by creating a one-off questionnaire on the menstrual cycle after downloading the app. The general somatic questions concerned symptoms as part of a psychosomatic syndrome or caused by the side effects of medication. Questions regarding psychological components were added. These questions concerned both negative and positive affect [20,35,36]. Social and contextual items were added because they could influence physical and emotional well-being and, therefore, the severity of the complaints [20,37]. Questions regarding food intake, use



XSL•FO
RenderX

of pain medication, and alcohol consumption were considered essential for influencing pain symptoms or general well-being.

## Phase III: Expert meeting

During the final expert meeting, 6 items were excluded and 1 item was added. The question "How many times did you wake up last night?" was excluded on the basis of relevance, as the quality of sleep and the reason for waking up had already been assessed. In the psychological items, the questions *I feel lonely* and *I feel insecure* were excluded to shorten the list. Furthermore, the ATLAS.ti analysis revealed that these items were mentioned less frequently by patients. Three questions with synonyms regarding energy level (feeling tired, feeling lifeless, and feeling energetic) were adapted to 1 question. The final questionnaire consisted of 51 items (Figure 1). The domains defined during the question selection phase were retained. The number of ESM items varied depending on the answers given by the patients. A morning questionnaire comprised a minimum of 4 and a maximum of 7 questions and included information about sleep and sexuality. Momentary assessments regarding the remaining domains comprised a minimum of 31 and a maximum of 42 items. Table 2 shows the number of questions per category. Two questions were added to a one-off questionnaire on the menstrual cycles of the patients.

**Table 2.** Number of experience sampling method (ESM) questions per category.

| Category | Maximum number of ESM questions |
|---|---|
| **One-off questionnaire** | |
| Menstrual cycle | 2 |
| **Morning questionnaire** | |
| Sleep | 4 |
| Sexuality | 3 |
| **Momentary assessment** | |
| Endometriosis-specific symptoms | 15 |
| General somatic symptoms | 7 |
| Mood and psychological factors | 7 |
| Social and contextual factors | 8 |
| Use of nutrition and medication | 5 |

## Phase IV: Development of a Smartphone App

The final questionnaire that was built into the smartphone app MEASuRE is listed in English in Multimedia Appendix 1. This questionnaire was originally created in Dutch and was officially translated by Medilingua translations; however, it has not yet been validated in English.

## Phase V: Pilot Study

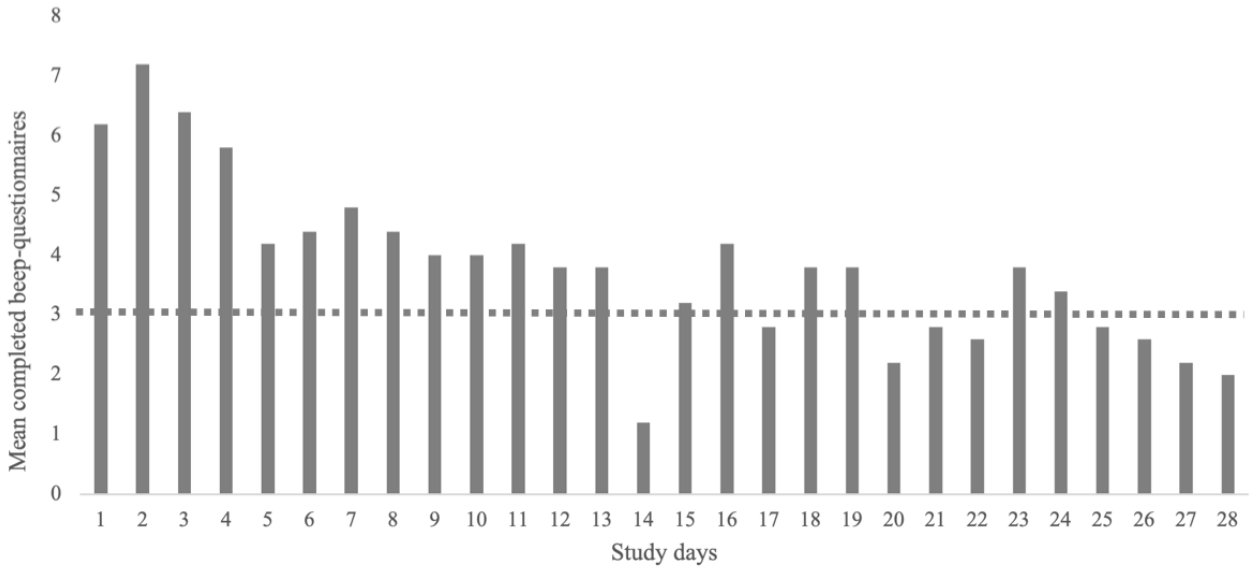
### *Feasibility and Compliance*

The characteristics of the women who participated in the pilot study are summarized in Table 1. The morning questionnaire took an average of 22 seconds to complete (range 11-44 seconds), and the momentary assessments took an average of 3 minutes and 2 seconds to complete (range of 72-255 seconds). The average completion rate for the morning questionnaires was 81% (23/28 study days). The average response rate for all momentary assessments was 37.86% (530/1400 questionnaires), with a range of 6.1% (17/280) to 56.1% (157/280) between patients. The average completion rate for a minimum of 3 questionnaires was 68% (19/28 study days). The response rate was highest during the first week of the pilot study, on average, 56% (39/70) of questionnaires, with a range of 21% (15/70) to 79% (55/70) between patients. The first week was the only week in which all participants completed at least 3 questionnaires on each study day. Figure 2 shows a histogram with the mean number of completed beep questionnaires (momentary assessments) per study day. In total, 0.79% (11/1400) of the momentary assessments were started but not completed.



XSL•FO
RenderX

**Figure 2.** Mean completed momentary assessments per study day. The horizontal dotted line represents the minimum of 3 completed questionnaires per day that is needed for a reliable analysis.




### *Interview*

During interviews, the patients noted that the app was easy to use and that the questions were clearly defined, although some suggestions were made for clarifications or answer options. Table 3 shows the results of the interviews with patients and includes advantages and limitations. Recommendations from these patients were also included, and their recommendations concerning the content were added to the final questionnaire (Multimedia Appendix 1). There were no particular questions that the patients did not want to answer, such as questions concerning sexuality.

**Table 3.** Advantages, limitations, and recommendations mentioned by the patients that completed the pilot study (n=5).

| Theme | Advantages | Limitations | Recommendations |
|---|---|---|---|
| Usability | • "Easy to use."<br>• "Completing a questionnaire was getting easier and faster over time." | • "Loading of the questionnaire was slow in case of a bad internet connection." | • Measure period for a maximum of 7 days:<br>  • "During a menstrual period" (n=1)<br>  • "Let the patient decide when to start measuring (when the complaints are highest)." (n=3)<br>• "Measure period for one month to be sure that one whole menstrual cycle is included. Use less questionnaires per day to improve compliance." (n=1) |
| Content | • "Very complete."<br>• "The questions are very clear."<br>• "Only by filling in my symptoms and activities at the same time makes me aware of a symptom pattern."<br>• "This is the only questionnaire I know that also contains bowel and bladder symptoms." | • See recommendations for missing items according to the pilot study participants. | • "The option I have pain while laying down is missing."<br>• "I am together with my pet."<br>• "The activity: taking care of my children/family."<br>• "Add LEFT (L) and RIGHT (R) to the abdominal pain figure." |
| Compliance | • "Timing ten times a day is good because then I don't feel guilty when I miss a questionnaire." | • "Work makes it difficult to complete the questionnaires."<br>• "Social activities make it difficult to complete the questionnaire."<br>• "Less motivation to fill in questionnaires when I don't have any somatic complaints."<br>• "Less motivation to fill in questionnaires when I am feeling down." | • "Completing 5-7 assessments per day is feasible." (n=3)<br>• "Completing 4-5 assessments per day is feasible." (n=2) |



XSL•FO
RenderX

## Discussion

### Overview

Following the development of an ESM tool specific to psychiatric conditions [20] and gastrointestinal and urological disorders, such as IBS, functional dyspepsia, or overactive bladder syndrome [22-26], we developed a modern assessment tool for patients with endometriosis. This new tool was developed according to the international guidelines on PROM development and comprised 5 phases: a selection of items on the basis of a literature review, a focus group study, expert meetings, the development of an electronic PROM using a smartphone app, and testing of the usability and feasibility with a pilot study. During interviews, the patients noted that the app was easy to use and that the questions were clearly defined. During our pilot study, only 0.79% (11/1400) of all momentary assessments were started but not completed, indicating that the assessments were easy to complete and not too time-consuming. However, completing up to 10 momentary assessments each day was considered time-consuming and caused response fatigue and noncompliance. During a study period of 28 days, most assessments were completed during the first week (39/70, 56%, in the first week vs on average 106/280, 38%, during the total study period). Compared with other ESM studies, this compliance rate is relatively low, as meta-analyses have shown completion rates of 82% to 85% [38,39]. However, comparing data with other ESM studies is difficult because the absence of methodological guidelines related to the use of this method has resulted in a large heterogeneity of designs [39], and compliance rates have not been reported in approximately half of the studies [40]. For better compliance, fewer study days, less assessments per day, and fewer items per assessment are advised [38,40]. In addition, as previous ESM studies recommend at least 3 completed questionnaires per day for a reliable analysis, which occurred consistently only during the first week of this study, we recommend using the ESM for a maximum of 7 days [41]. However, as endometriosis can fluctuate during the menstrual cycle, assessing patients for 4 weeks could add valuable information and might be considered with fewer assessments per day.

### Strengths and Limitations of the ESM

The ESM has several advantages over traditionally used assessment tools, including the ability to evaluate the temporal relationship between variables, high ecological validity, and highly detailed information on the experience of the subject. Furthermore, the ESM allows for a prospective, individualized within - person approach to symptoms and symptom formation and to treatment outcome, which contrasts with the *average patient* approach of traditional evidence-based practice [37,41]. Self-reports across multiple days and among various participants provide profound and comprehensive insights into the disease course and treatment efficacy. On the basis of this, the ESM may also provide clues for behavioral interventions, adding value to fragmented monodisciplinary treatment, which remains refractory to responsiveness.

A limitation of the ESM is that it is perceived as time-consuming and requires considerable motivation on the part of the patient. Therefore, assessments are ideally kept as brief as possible. Furthermore, assessments several days in a row could encourage rumination. Thus, on the basis of the recommendations of previous studies using ESM, we suggest limiting the assessment period to 7 days and adding items concerning positive affect [20,35,36]. Another concern is selection bias. Not all patients are willing to participate or comply with study protocols using ESM, and participation could be affected by motivation for change in treatment. However, previous research has shown that this method is feasible for a wide variety of patients [42,43].

### Strengths and Limitations of PROM Development

Given that the questions in this new tool are derived from validated questionnaires, this ESM tool designed for use in patients with endometriosis is comparable to validated retrospective PROMs. The use of patient focus groups according to the international guidelines on PROM development strengthens the validity of the questionnaire. A limitation of our focus group study was the limited number of patients who participated. Although 19 patients agreed to participate, only 14 were included in the 3 focus groups. Ideally, 6 to 10 participants were scheduled for each focus group. Most importantly, saturation of input was reached. During the pilot study, a few recommendations were made regarding the content, and these were added to the final questionnaire (Multimedia Appendix 1).

### Future Study Perspectives

This paper comprises the development (part I) of a new PROM for women with endometriosis, with the ability to assess symptoms in real time. The validation stage (part II) will involve testing the psychometric properties of this newly developed tool. A 7-day validation study will be conducted to assess content validity and to investigate the association with potential triggers of physical symptoms, such as psychological, social, and contextual factors. In the planned validation study, 25 patients with endometriosis with CPP at least 1 day per week on average will be included. By letting patients start measuring at random moments, we expect to collect enough data from different menstrual cycle phases and that there will be sufficient data after the use of ESM in 7 consecutive days. Data from this newly developed ESM tool will be compared with frequently used validated (retrospective) outcome measures such as the EHP-30 questionnaire and end-of-day and end-of-week retrospective pain scores. After validation and testing, future studies will be planned to evaluate the use of this tool in a clinical setting in order to propose a personalized treatment plan.

In conclusion, in agreement with the international guidelines, we developed a PROM for real-time symptom assessment in women with endometriosis. This new electronic tool consists of a morning questionnaire and momentary assessments with questions regarding physical, mental, sexual, and social well-being. This tool was considered easy to use and may help overcome the limitations of existing retrospective questionnaires. To minimize noncompliance, it is advised to use this tool for a maximum of 7 days.



XSL•FO
RenderX

## Conflicts of Interest

GVK performs clinical trials with Medtronic, Boston Scientific, and Astellas and is a consultant to Medtronic, Boston Scientific, and Solace therapeutics. None of these conflicts are relevant to this work.

Multimedia Appendix 1
Set of questions for the endometriosis-specific experience sampling method–patient-reported outcome measure after focus groups, expert meetings, and pilot study.
[DOCX File , 117 KB - formative_v5i12e28782_app1.docx ]

## References


1. Kennedy S, Bergqvist A, Chapron C, D'Hooghe T, Dunselman G, Greb R, ESHRE Special Interest Group for EndometriosisEndometrium Guideline Development Group. ESHRE guideline for the diagnosis and treatment of endometriosis. Hum Reprod 2005 Oct;20(10):2698-2704. [doi: 10.1093/humrep/dei135] [Medline: 15980014]
2. Dunselman GA, Vermeulen N, Becker C, Calhaz-Jorge C, D'Hooghe T, De Bie B, European Society of Human ReproductionEmbryology. ESHRE guideline: management of women with endometriosis. Hum Reprod 2014 Mar;29(3):400-412. [doi: 10.1093/humrep/det457] [Medline: 24435778]
3. Shafrir A, Farland L, Shah D, Harris H, Kvaskoff M, Zondervan K, et al. Risk for and consequences of endometriosis: a critical epidemiologic review. Best Pract Res Clin Obstet Gynaecol 2018 Aug;51:1-15. [doi: 10.1016/j.bpobgyn.2018.06.001] [Medline: 30017581]
4. Bellelis P, Dias JJ, Podgaec S, Gonzales M, Baracat EC, Abrão MS. Epidemiological and clinical aspects of pelvic endometriosis–a case series. Rev Assoc Med Bras (1992) 2010;56(4):467-471 [FREE Full text] [doi: 10.1590/s0104-42302010000400022] [Medline: 20835646]
5. Davis GD, Thillet E, Lindemann J. Clinical characteristics of adolescent endometriosis. J Adolesc Health 1993 Jul;14(5):362-368. [doi: 10.1016/s1054-139x(08)80008-0] [Medline: 8399247]
6. Chaman-Ara K, Bahrami M, Moosazadeh M. Quality of life in women with endometriosis: a systematic review and meta-analysis. World Cancer Res J 2017;4(1):e839. [doi: 10.32113/wcrj_20173_839]
7. Culley L, Law C, Hudson N, Denny E, Mitchell H, Baumgarten M, et al. The social and psychological impact of endometriosis on women's lives: a critical narrative review. Hum Reprod Update 2013;19(6):625-639. [doi: 10.1093/humupd/dmt027] [Medline: 23884896]
8. Nnoaham KE, Hummelshoj L, Webster P, d'Hooghe T, de Cicco Nardone F, de Cicco Nardone C, World Endometriosis Research Foundation Global Study of Women's Health consortium. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. Fertil Steril 2011 Aug;96(2):366-73.e8 [FREE Full text] [doi: 10.1016/j.fertnstert.2011.05.090] [Medline: 21718982]
9. Simoens S, Dunselman G, Dirksen C, Hummelshoj L, Bokor A, Brandes I, et al. The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. Hum Reprod 2012 May;27(5):1292-1299. [doi: 10.1093/humrep/des073] [Medline: 22422778]
10. Rush G, Misajon R, Hunter JA, Gardner J, O'Brien KS. The relationship between endometriosis-related pelvic pain and symptom frequency, and subjective wellbeing. Health Qual Life Outcomes 2019 Jul 16;17(1):123 [FREE Full text] [doi: 10.1186/s12955-019-1185-y] [Medline: 31311560]
11. Becker CM, Gattrell WT, Gude K, Singh SS. Reevaluating response and failure of medical treatment of endometriosis: a systematic review. Fertil Steril 2017 Jul;108(1):125-136 [FREE Full text] [doi: 10.1016/j.fertnstert.2017.05.004] [Medline: 28668150]
12. Vercellini P, Fedele L, Aimi G, Pietropaolo G, Consonni D, Crosignani PG. Association between endometriosis stage, lesion type, patient characteristics and severity of pelvic pain symptoms: a multivariate analysis of over 1000 patients. Hum Reprod 2007 Jan;22(1):266-271. [doi: 10.1093/humrep/del339] [Medline: 16936305]
13. Vitale SG, La Rosa VL, Rapisarda AM, Laganà AS. Impact of endometriosis on quality of life and psychological well-being. J Psychosom Obstet Gynaecol 2017 Dec;38(4):317-319. [doi: 10.1080/0167482X.2016.1244185] [Medline: 27750472]
14. Laganà AS, La Rosa VL, Rapisarda AM, Valenti G, Sapia F, Chiofalo B, et al. Anxiety and depression in patients with endometriosis: impact and management challenges. Int J Womens Health 2017 May 16;9:323-330 [FREE Full text] [doi: 10.2147/IJWH.S119729] [Medline: 28553145]
15. Vitale SG, Petrosino B, La Rosa VL, Rapisarda AM, Laganà AS. A systematic review of the association between psychiatric disturbances and endometriosis. J Obstet Gynaecol Can 2016 Dec;38(12):1079-1080. [doi: 10.1016/j.jogc.2016.09.008] [Medline: 27986180]
16. Zondervan KT, Becker CM, Missmer SA. Endometriosis. N Engl J Med 2020 Mar 26;382(13):1244-1256. [doi: 10.1056/NEJMra1810764] [Medline: 32212520]




XSL•FO
RenderX

17. Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. Pain 2009 Dec;146(3):238-244. [doi: 10.1016/j.pain.2009.08.019] [Medline: 19836888]
18. Vincent K, Kennedy S, Stratton P. Pain scoring in endometriosis: entry criteria and outcome measures for clinical trials. Report from the Art and Science of Endometriosis meeting. Fertil Steril 2010 Jan;93(1):62-67 [FREE Full text] [doi: 10.1016/j.fertnstert.2008.09.056] [Medline: 18990378]
19. Jones G, Jenkinson C, Kennedy S. Evaluating the responsiveness of the Endometriosis Health Profile Questionnaire: the EHP-30. Qual Life Res 2004 Apr;13(3):705-713. [doi: 10.1023/B:QURE.0000021316.79349.af] [Medline: 15130032]
20. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, ESM-MERGE Investigators PhD, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 Jun;34(6):481-493. [doi: 10.1002/da.22647] [Medline: 28544391]
21. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
22. Mujagic Z, Leue C, Vork L, Lousberg R, Jonkers DM, Keszthelyi D, et al. The Experience Sampling Method--a new digital tool for momentary symptom assessment in IBS: an exploratory study. Neurogastroenterol Motil 2015 Sep;27(9):1295-1302. [doi: 10.1111/nmo.12624] [Medline: 26100684]
23. Vork L, Keszthelyi D, Mujagic Z, Kruimel JW, Leue C, Pontén I, et al. Development, content validity, and cross-cultural adaptation of a patient-reported outcome measure for real-time symptom assessment in irritable bowel syndrome. Neurogastroenterol Motil 2018 Mar;30(3). [doi: 10.1111/nmo.13244] [Medline: 29106029]
24. Vork L, Mujagic Z, Drukker M, Keszthelyi D, Conchillo JM, Hesselink MA, et al. The Experience Sampling Method-Evaluation of treatment effect of escitalopram in IBS with comorbid panic disorder. Neurogastroenterol Motil 2019 Jan;31(1):e13515 [FREE Full text] [doi: 10.1111/nmo.13515] [Medline: 30460734]
25. Smeets FG, Keszthelyi D, Vork L, Tack J, Talley NJ, Simren M, et al. Development of a real-time patient-reported outcome measure for symptom assessment in patients with functional dyspepsia using the experience sampling method. Neurogastroenterol Motil 2019 Feb;31(2):e13496. [doi: 10.1111/nmo.13496] [Medline: 30393939]
26. Herrewegh A, Vork L, Eurelings E, Leue C, Kruimel J, van Koeveringe G, et al. The development of a patient-reported outcome measure for real-time symptom assessment in a population with functional urologic complaints-A focus group study. Neurourol Urodyn 2018 Nov;37(8):2893-2903. [doi: 10.1002/nau.23808] [Medline: 30187953]
27. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006 Oct 11;4:79 [FREE Full text] [doi: 10.1186/1477-7525-4-79] [Medline: 17034633]
28. Morgan D. The Focus Group Guidebook. Portland: Sage Publications; 1998.
29. Myin-Germeys I, van Os J, Schwartz JE, Stone AA, Delespaul PA. Emotional reactivity to daily life stress in psychosis. Arch Gen Psychiatry 2001 Dec;58(12):1137-1144. [doi: 10.1001/archpsyc.58.12.1137] [Medline: 11735842]
30. Husky MM, Grondin OS, Swendsen JD. The relation between social behavior and negative affect in psychosis-prone individuals: an experience sampling investigation. Eur Psychiatry 2004 Feb;19(1):1-7. [doi: 10.1016/j.eurpsy.2003.09.006] [Medline: 14969774]
31. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. Scand J Gastroenterol 1997 Sep;32(9):920-924. [doi: 10.3109/00365529709011203] [Medline: 9299672]
32. Bowden A, Sabounjian L, Sandage B. Psychometric validation of an urgency severity scale (IUSS) for patients with overactive bladder. ics.org. URL: https://www.ics.org/Abstracts/Publish/41/000119.pdf [accessed 2021-11-07]
33. Barbara G, Facchin F, Buggio L, Somigliana E, Berlanda N, Kustermann A, et al. What is known and unknown about the association between endometriosis and sexual functioning: a systematic review of the literature. Reprod Sci 2017 Dec;24(12):1566-1576. [doi: 10.1177/1933719117707054] [Medline: 28558521]
34. Jones G, Jenkinson C, Kennedy S. The Endometriosis Health Profile User Manual: user manual for the EHP-30 and the EHP-5. Nuffield Department of Obstetrics & Gynaecology & Health Services Research Unit University of Oxford. 2001. URL: https://innovation.ox.ac.uk/wp-content/uploads/2014/09/The-Endometriosis-Health-Profile-User-Manual.docx [accessed 2021-11-07]
35. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019 Apr 09;365:l1476 [FREE Full text] [doi: 10.1136/bmj.l1476] [Medline: 30967483]
36. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
37. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. Evid Based Ment Health 2016 Aug;19(3):86-89 [FREE Full text] [doi: 10.1136/ebmental-2016-102418] [Medline: 27443678]



XSL•FO
RenderX

38. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. J Med Internet Res 2019 Feb 05;21(2):e11398 [FREE Full text] [doi: 10.2196/11398] [Medline: 30720437]
39. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations. J Med Internet Res 2019 Dec 06;21(12):e14475 [FREE Full text] [doi: 10.2196/14475] [Medline: 31808748]
40. Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. J Med Internet Res 2021 Mar 03;23(3):e17023 [FREE Full text] [doi: 10.2196/17023] [Medline: 33656451]
41. Palmier-Claus J, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul PA, et al. Experience sampling research in individuals with mental illness: reflections and guidance. Acta Psychiatr Scand 2011 Jan;123(1):12-20. [doi: 10.1111/j.1600-0447.2010.01596.x] [Medline: 20712828]
42. van Berkel N, Ferreira D, Kostakos V. The experience sampling method on mobile devices. ACM Comput Surv 2018 Jan 12;50(6):1-40. [doi: 10.1145/3123988]
43. Verhagen SJ, Berben JA, Leue C, Marsman A, Delespaul PA, van Os J, et al. Demonstrating the reliability of transdiagnostic mHealth Routine Outcome Monitoring in mental health services using experience sampling technology. PLoS One 2017 Oct 12;12(10):e0186294 [FREE Full text] [doi: 10.1371/journal.pone.0186294] [Medline: 29023568]

## Abbreviations

**CPP:** chronic pelvic pain
**EHP-30:** Endometriosis Health Profile-30
**ESM:** experience sampling method
**IBS:** irritable bowel syndrome
**MEASuRE:** Maastricht Electronic Abdominal Symptom Reporting
**MUMC+:** Maastricht University Medical Centre
**PROM:** patient-reported outcome measure

*Edited by G Eysenbach; submitted 15.03.21; peer-reviewed by MS Qureshi, PC Masella; comments to author 29.07.21; revised version received 22.09.21; accepted 15.10.21; published 03.12.21.*

*Please cite as:*
*van Barneveld E, Lim A, van Hanegem N, Vork L, Herrewegh A, van Poll M, Manders J, van Osch F, Spaans W, van Koeveringe G, Vrijens D, Kruimel J, Bongers M, Leue C*
*Patient-Reported Outcome Measure for Real-time Symptom Assessment in Women With Endometriosis: Focus Group Study*
*JMIR Form Res 2021;5(12):e28782*
*URL: https://formative.jmir.org/2021/12/e28782*
*doi:10.2196/28782*
*PMID:34870608*

©Esther van Barneveld, Arianne Lim, Nehalennia van Hanegem, Lisa Vork, Alexandra Herrewegh, Mikal van Poll, Jessica Manders, Frits van Osch, Wilbert Spaans, Gommert van Koeveringe, Desiree Vrijens, Joanna Kruimel, Marlies Bongers, Carsten Leue. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Assessing the Views of Professionals, Patients, and Care Partners Concerning the Use of Computer Tools in Memory Clinics: International Survey Study

Aniek M van Gils[1], MSc, MD; Leonie NC Visser[1,2], PhD; Heleen MA Hendriksen[1], MSc; Jean Georges[3], AB; Majon Muller[4], MD, Prof Dr; Femke H Bouwman[1], MD, PhD; Wiesje M van der Flier[1,5], Prof Dr; Hanneke FM Rhodius-Meester[1,4], MD, PhD

[1]Department of Neurology, Alzheimer Center Amsterdam, Amsterdam Neuroscience, Amsterdam UMC, Location VUmc, Amsterdam, Netherlands

[2]Department of Neurobiology, Care Sciences and Society, Division of Clinical Geriatrics, Center for Alzheimer Research, Karolinska Institutet, Stockholm, Sweden

[3]Alzheimer Europe, Luxembourg, Luxembourg

[4]Department of Internal Medicine, Geriatric Medicine Section, Amsterdam Cardiovascular Sciences Institute, Amsterdam UMC, Location VUmc, Amsterdam, Netherlands

[5]Department of Epidemiology and Data Science, Amsterdam UMC, Location VUmc, Amsterdam, Netherlands

**Corresponding Author:**
Aniek M van Gils, MSc, MD
Department of Neurology, Alzheimer Center Amsterdam, Amsterdam Neuroscience
Amsterdam UMC, Location VUmc
De Boelelaan 1118
Amsterdam, 1081 HZ
Netherlands
Phone: 31 204440685
Email: a.vangils@amsterdamumc.nl

## Abstract

**Background:** Computer tools based on artificial intelligence could aid clinicians in memory clinics in several ways, such as by supporting diagnostic decision-making, web-based cognitive testing, and the communication of diagnosis and prognosis.

**Objective:** This study aims to identify the preferences as well as the main barriers and facilitators related to using computer tools in memory clinics for all end users, that is, clinicians, patients, and care partners.

**Methods:** Between July and October 2020, we sent out invitations to a web-based survey to clinicians using the European Alzheimer's Disease Centers network and the Dutch Memory Clinic network, and 109 clinicians participated (mean age 45 years, SD 10; 53/109, 48.6% female). A second survey was created for patients and care partners. They were invited via Alzheimer Europe, Alzheimer's Society United Kingdom, Amsterdam Dementia Cohort, and Amsterdam Aging Cohort. A total of 50 patients with subjective cognitive decline, mild cognitive impairment, or dementia (mean age 73 years, SD 8; 17/34, 34% female) and 46 care partners (mean age 65 years, SD 12; 25/54, 54% female) participated in this survey.

**Results:** Most clinicians reported a willingness to use diagnostic (88/109, 80.7%) and prognostic (83/109, 76.1%) computer tools. User-friendliness (71/109, 65.1%); Likert scale mean 4.5, SD 0.7), and increasing diagnostic accuracy (76/109, 69.7%; mean 4.3, SD 0.7) were reported as the main factors stimulating the adoption of a tool. Tools should also save time and provide clear information on reliability and validity. Inadequate integration with electronic patient records (46/109, 42.2%; mean 3.8, SD 1.0) and fear of losing important clinical information (48/109, 44%; mean 3.7, SD 1.2) were most frequently indicated as barriers. Patients and care partners were equally positive about the use of computer tools by clinicians, both for diagnosis (69/96, 72%) and prognosis (73/96, 76%). In addition, most of them thought favorably regarding the possibility of using the tools themselves.

**Conclusions:** This study showed that computer tools in memory clinics are positively valued by most end users. For further development and implementation, it is essential to overcome the technical and practical barriers of a tool while paying utmost attention to its reliability and validity.

***(JMIR Form Res 2021;5(12):e31053)*** doi:10.2196/31053



XSL•FO
RenderX

**KEYWORDS**

artificial intelligence; clinical decision support systems; dementia; diagnostic testing; diagnosis; prognosis; communication

# Introduction

## Background

Dementia is a major health problem worldwide, with its prevalence expected to rise to 75 million patients in 2050 [1]. A timely and accurate diagnosis is essential for providing adequate care and appropriate treatment [2]. Diagnosing Alzheimer disease dementia or another type of dementia can be challenging, as clinical presentations overlap and multiple pathologies often co-occur [3,4]. Furthermore, increasing biomarker availability creates the possibility of diagnosing the early stages of Alzheimer disease before the onset of dementia and paves the way for individual dementia risk estimation [5,6].

With the availability of many different diagnostic tests, clinicians have the difficult task of combining and interpreting all the test results to come to an accurate diagnosis and prognosis [7-10] and communicating these results to patients and care partners [5,11,12]. Despite the increasing number of diagnostic tools, uncertainty in the diagnosis and prognosis of dementia remains, complicating the process of clearly explaining the test results [13]. Recent research has shown that most patients and care partners greatly value precise information on diagnosis and prognosis [12,14]. However, clinicians are often reluctant to address these topics during consultations, leaving these informational needs unmet [12,15].

Currently, artificial intelligence solutions are rapidly being developed and can aid clinicians in addressing these challenges in several ways. Artificial intelligence–based computer tools for dementia diagnosis and prognosis have demonstrated diagnostic accuracy equal to that of clinicians' performance. These tools support individual risk estimation and increase clinicians' confidence in diagnosis and prognosis [16-19]. Web-based cognitive test tools have shown promising results, enabling cost-effective testing [20-22]. From other medical fields, such as oncology, we know that computer-based tools can also support the communication process, for example, by engaging patients and their families more actively in the diagnostic decisions or by supporting clinicians in the clear communication of results [23,24]. To date, the actual implementation of such computer tools in memory clinic practice has been limited [25,26].

## Barriers and Facilitators

Several barriers to the acceptance and implementation of tools have been identified in different health care areas. The main concerns regarding computer tools are related to the physician–patient relationship: the fear of interfering with this relationship when using a tool and affecting patient communication. Furthermore, clinicians fear the disturbance of clinical work and the loss of clinical autonomy when using a tool. In addition, a time-consuming tool, a tool that does not fit into the workflow, complexity of a tool, and computer literacy have been frequently mentioned as barriers. On the other hand, good training before the use of a tool, user-friendliness, relevancy, transparency, and reliability are stimulating factors in the use of a computer tool in clinical practice [27-33].

It is not known if the same barriers and facilitators apply to computer tools in memory clinics. The nature of the patient population—older adults with cognitive decline—and the vast number of diagnostic tests involved in the diagnostic process might lead to a different set of relevant barriers and facilitators. In addition, patients' and care partners' opinions regarding the use of computer tools by their clinicians might be a potential barrier to or facilitator of clinicians' use of a tool.

## Objective

Therefore, this study aims to understand preferences and identify the main barriers to and facilitators of using computer tools in the dementia workup from the perspectives of clinicians, patients, and care partners, that is, the end users.

# Methods

## Design

We conducted 2 surveys, 1 for clinicians and 1 directed at patients and care partners, both in the fall of 2020. In addition, to aid the interpretation of the survey results, we conducted an interactive panel session with clinicians. The study was approved by the medical ethical committee of the Amsterdam University Medical Center, Vrije Universiteit Medical Center, Amsterdam. All participants provided digitally informed consent.

## Survey for Clinicians

### *Participants*

Between July and October 2020, we invited clinicians from memory clinics in Europe via the European Alzheimer's Disease Consortium and the Dutch Memory Clinic network (Nederlands Geheugenpoli Netwerk [NGN]) through emails that contained a link to participate in the web-based survey. Furthermore, we invited all participants during the annual NGN conference to share their thoughts in our web-based interactive panel session.

### *Survey*

The survey was created in the web-based survey tool Survalyzer (Survalyzer AG) [34] and translated into Dutch and English. The survey was adaptive; that is, certain questions were only conditionally displayed based on responses to other items. Furthermore, participants could scroll through the survey to edit their answers. The survey comprised 3 parts. In the first part, we collected background information (eg, age, gender, profession, and specialization). In the second part, we used a funneled method to examine the current opinions on computer tools and identify the barriers and facilitators. First, we asked if the clinician would be willing to use computer tools in general, after which we asked them to explain their opinions in an open-ended question. Subsequently, we provided them with a list of barriers and facilitators compiled from barriers and facilitators known from the existing literature [27,28,32], and we asked them which factors would stimulate or discourage



XSL•FO
RenderX

them from using a tool. Clinicians could complement this list with their own perspectives. We then asked them to rate the importance of these factors using the Likert scale (1=very unimportant, 2=unimportant, 3=neutral, 4=important, and 5=very important). Finally, we asked the participants how likely they were to use *diagnostic* and *prognostic* tools. In the third part, we explored clinicians' opinions on additional computer tools, that is, web-based cognitive testing, communication support, and communication skills training.

### *Interactive Panel Session*

During the NGN annual conference (held on the web on November 10, 2020), 2 authors (HFMRM and LNCV) presented the preliminary results of the survey. To help interpret these results, they asked all conference participants several in-depth questions using Mentimeter [35]. These questions were related to the importance of several factors that stimulate their trust in a tool; factors that would convince them of the usability, reliability, and validity of a tool; and the primary outcome measures of a tool.

## Survey for Patients and Care Partners

### *Participants*

Between July and October 2020, we invited a mixed memory clinic population comprising patients with subjective memory complaints (subjective cognitive decline [SCD]), mild cognitive impairment, and dementia and care partners to participate in the web-based survey. To be included, patients had to be able to understand the questionnaire in Dutch or English. In this study, *care partner* refers either to an informal caregiver or a close relative or friend of the patient who provides either or both emotional and practical support. A general link to the survey was sent via a newsletter and social media to the members (patients and care partners) of Alzheimer Europe and directly to the members of the Alzheimer's Society United Kingdom. Next, we invited both patients and care partners from the Amsterdam Dementia Cohort of the Alzheimer Center Amsterdam [36,37] and the Amsterdam Aging Cohort of the outpatient geriatric clinic of the Amsterdam University Medical Center (Amsterdam University Medical Centers) [38]. Patients and care partners were approached and informed by phone or email, and when they confirmed their participation, they were sent a personalized link to the web-based survey. The survey was adaptive to reduce the number of questions. Furthermore, participants could scroll through the survey to edit their answers.

### *Survey*

A total of 2 versions of the survey were created, 1 directed at patients and 1 at care partners, both comprising 3 parts. In part 1, we collected background information regarding the participants (eg, age, gender, and diagnosis). In part 2, we asked for their opinion on clinicians' use of (1) a computer tool that analyzes the results of the diagnostic tests (*diagnostic tool*), (2) a computer tool to help predict the course of their symptoms (*prognostic tool*), and (3) a tool to help communicate the test results in *day-to-day* language with the patient (*communication tool*). We adjusted the predefined list of barriers to and facilitators of using computer tools to the patient and care partner perspectives. We provided participants with this list and asked the extent (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree) to which the different items applied to them. In the last part, we asked their opinion on computer tools directed at patients and care partners, that is, web-based cognitive testing and tools that could support and empower them in their communication with the clinicians. The survey was piloted in a test panel of 3 patients (2 with SCD and 1 with dementia) and 1 care partner of a person with dementia.

## Analysis

Completion of the survey was enforced using mandatory questions. Proceeding with the survey was not possible when a question was unanswered. Furthermore, it was not possible to complete the survey more than once, as Survalyzer solely allowed unique visitors. Only completed surveys were analyzed. We analyzed participant characteristics and survey outcomes using descriptive statistics. Chi-square tests were used to compare answers between patients and care partners. For clinicians, we compared answers between groups based on age, sex, profession, and specialization. When using the 5-point Likert scale, the mean Likert scale scores were calculated per item. Frequencies were calculated for all the barriers and facilitators from the predefined list. We combined the frequencies with mean Likert scale scores to define the most important barriers and facilitators (eg, the item with the highest frequency combined with the highest mean Likert scale score was regarded as the most important). SPSS, version 22.0 (IBM Corporation) was used to analyze the quantitative data. $P$ values <.05 were considered significant.

The answers to the open-ended questions were analyzed in MAXQDA software (VERBI Software) [39] using a process of deductive thematic content analysis [40,41]. A total of 2 authors, AMVG (physician) and HMAH (neuropsychologist), independently generated the initial thematic codes based on the existing literature and data. Subsequently, 1 author (AMVG) generated a thematic framework and used this framework to code all the given answers. The codes were then sorted into broad categories.

# Results

## Demographics

Sample descriptions have been presented in Table 1 for clinicians and Table 2 for patients and care partners. Clinicians were, on average, aged 45 (SD 11) years and had 16 (SD 13) years of experience. Most participating clinicians were medical specialists working in neurology (60/109, 55%) or internal or clinical geriatric medicine (33/109, 30.3%). Patients were in general older (mean age 73 years, SD 8) than care partners (mean age 65 years, SD 12), who were mostly a partner or spouse (33/46, 72%) or a granddaughter, daughter, grandson, and son (in-law; 12/46, 26%). Participating patients were most often diagnosed with SCD (21/50, 42%), whereas participating care partners were mostly those of patients with dementia (36/46, 78%).



XSL•FO
RenderX

**Table 1.** Sample demographics of clinicians participating in the web-based survey and interactive panel session (N=294).

| Characteristics | Web-based survey (n=109) | Interactive panel session[a] (n=184) |
|---|---|---|
| Age (years), mean (SD) | 45 (11) | 43 (11) |
| Sex (female), n (%) | 53 (48.6) | 98 (85.9) |
| **Cohort, n (%)** | | |
| European Alzheimer's Disease Consortium | 53 (48.6) | N/A[b] |
| Dutch Memory Clinic network | 56 (51.4) | N/A |
| **Profession, n (%)** | | |
| MD[c], specialist | 87 (79.8) | 60 (54.5) |
| MD, specialist training or not in specialist training | 12 (10.9) | 1 (0.9) |
| Physician assistant, nurse specialist, or specialized nurse | 3 (2.8) | 23 (20.9) |
| Neuropsychologist or psychologist | 6 (5.5) | 16 (114.5) |
| Other | 1 (0.9) | 10 (9.1) |
| Experience[d] (years), mean (SD) | 16 (13) | N/A |
| **Specialization[e], n (%)** | | |
| Neurology | 60 (50.4) | N/A |
| Clinical geriatric or internal medicine | 33 (30.3) | N/A |
| Nursing home physician or general practitioner | 2 (1.8) | N/A |
| Psychiatry | 9 (8.3) | N/A |
| Other | 9 (8.3) | N/A |
| **Institution[f], n (%)** | | |
| Academic or university hospital | 68 (62.4) | N/A |
| Nonacademic teaching hospital | 32 (29.4) | N/A |
| Nonteaching hospital | 8 (7.3) | N/A |
| Mental health service | 2 (1.8) | N/A |
| Other | 3 (2.8) | N/A |

[a]Owing to the hybrid conference setting, not every description was available for all participants. For sex, n=114 participants replied to the question. For profession, n=110 participants replied to the question.

[b]N/A: not applicable.

[c]MD: medical doctor.

[d]Only applicable for medical specialists.

[e]Some clinicians had ≥1 specialization.

[f]Some clinicians worked in ≥1 institution.



XSL•FO
RenderX

**Table 2.** Sample demographics of patients and care partners participating in the web-based survey (N=96).

| Characteristics | Patients[a] (n=50) | Care partners (n=46) |
|---|---|---|
| Age (years), mean (SD) | 73 (8) | 65 (12) |
| Sex (female), n (%) | 17 (34) | 25 (54) |
| **Cohort, n (%)** | | |
| Alzheimer Europe or Alzheimer's Society United Kingdom | 2 (4) | 14 (30) |
| Amsterdam dementia cohort | 25 (50) | 27 (52) |
| Amsterdam aging cohort | 23 (46) | 5 (18) |
| **Diagnosis[b], n (%)** | | |
| SCD[c] | 21 (42) | 2 (4) |
| MCI[d] | 16 (32) | 8 (17) |
| Dementia | 13 (26) | 36 (78) |
| **Education[e], n (%)** | | |
| Low | 1 (2) | 1 (2) |
| Middle | 22 (45) | 16 (36) |
| High | 26 (53) | 27 (61) |

[a]Of these 50 patients, 20 (40%) completed the survey together with their care partner.

[b]For the care partners, the numbers represent the diagnosis of their loved ones.

[c]SCD: subjective cognitive decline.

[d]MCI: mild cognitive impairment.

[e]According to the Dutch Verhage scale (low 1-3; middle 5; high 6-7).

## Survey of Clinicians

### *Opinions on the Use of Computer Tools*

In response to whether they would be willing to use computer tools in their daily clinical practice, 51.4% (56/109) of clinicians said they would probably use a diagnostic tool, and 29.4% (32/109) said they would certainly use a diagnostic tool. Furthermore, 7.3% (8/109) said they would be unlikely to use a tool, 11.9% (13/109) answered neutrally, and none of the clinicians reported that they did not want to use a tool. The results were similar for prognostic tools; of the 109 clinicians, 53 (48.6%) said they would probably use a prognostic tool, and 30 (27.6%) said they would certainly use a prognostic tool. Furthermore, 2.8% (3/109) were unlikely to use a prognostic tool, and 0.9% (1/109) would certainly not use a prognostic tool. The remaining participants responded as neutral (22/109, 20.2%). We found no differences in both diagnostic and prognostic tools based on age ($P$=.20 and $P$=.49, respectively), sex ($P$=.14 and $P$=.73, respectively), or profession ($P$=.61 and $P$=.98, respectively). We found that neurologists indicated more willingness to use prognostic tools ($P$=.04) than clinicians from other specializations.

Content analysis of clinicians' explanations of their opinion on the use of computer tools resulted in 6 main topics: *support*, *clinical expertise*, *efficiency*, *accuracy*, *clinician–patient relationship*, and *care of the future*. Each of the topics has been described in Table 3, and illustrative quotes have been provided.



XSL•FO
RenderX

**Table 3.** Clinicians' opinions on the use of computer tools in memory clinics.

| Topics | Description | | Quotes |
|---|---|---|---|
| | Facilitating factors | Hindering factors | |
| Support | Support to the diagnostic process from screening or prescreening to follow up, support data storage, support research purposes | Not applicable for specific patient populations | • "[...] I would welcome a tool that would be implemented with the available clinical data and help reach a diagnosis (ie, considering the neurocognitive and neuroimaging data, in patients with such profile, expert diagnosis would be...with a probability of...%—which could be increased by the use of...biomarker)." [Male, 42 years, MD[b], physician working in neurology] |
| Clinical expertise | Complementary to clinical expertise (eg, an aid for complex cases or with interpretation of test results) and contributory to evidence-based medicine [42] | Tools should not be a replacement for clinical expertise | • "Computer tools and AI might be a way to have an evidence-based standard procedure in addition to my own long time clinical experience." [Female, 59 years, MD, geriatrician];<br>• "[...] I consider the clinical view as most important. A computer tool cannot (partly) replace this." [Female, 38 years, MD, geriatrician] |
| Efficiency | The ability to standardize the diagnostic process, if easy to use, if connecting with electronic patient file, and time-efficiency | A tool not connected to the electronic patient file, information technology issues | • "A quick and useful way to get practical answers on the workplace." [Male, 62 years, MD, neurologist];<br>• "For tools that are not implemented in the electronic patient file I foresee barriers in the implementation." [Female, 45 years, MD, geriatrician] |
| Accuracy | Computer tools could help in making a more accurate diagnosis, providing additional objective information, and overcoming human errors | Tools might generate results of no use and fear of loss of important clinical information | • "Sometimes we can be influenced by the patient we have in front of us. We can diagnose them too easily or consider them as (sub)normal because their general behavior makes us think so. A computer could be more objective than we are in some cases." [Male, 26 years, MD, neurology resident];<br>• "[...] I am afraid that there will be an outcome that is of no use for me, such as 64% chance of Alzheimer's disease." [Female, 38 years, MD, geriatrician] |
| Clinician–patient relationship | Improving patient communication | A tool might have a negative impact on the relationship between clinicians and patients | • "[...] It facilitates the communication to the patient." [Female, 49 years, MD, neurologist];<br>• "Patients also come for attention and care, which they get less if we look at the screen more often." [Male, 32 years, MD, physician working in neurology] |
| Care of the future | The use of tools is considered part of the care of the future | N/A[a] | • "AI and big data are the future, they make the invisible visible [...]." [Male, 33 years, MD, internal (geriatric) medicine resident] |

[a]N/A: not applicable.

[b]MD: medical doctor.

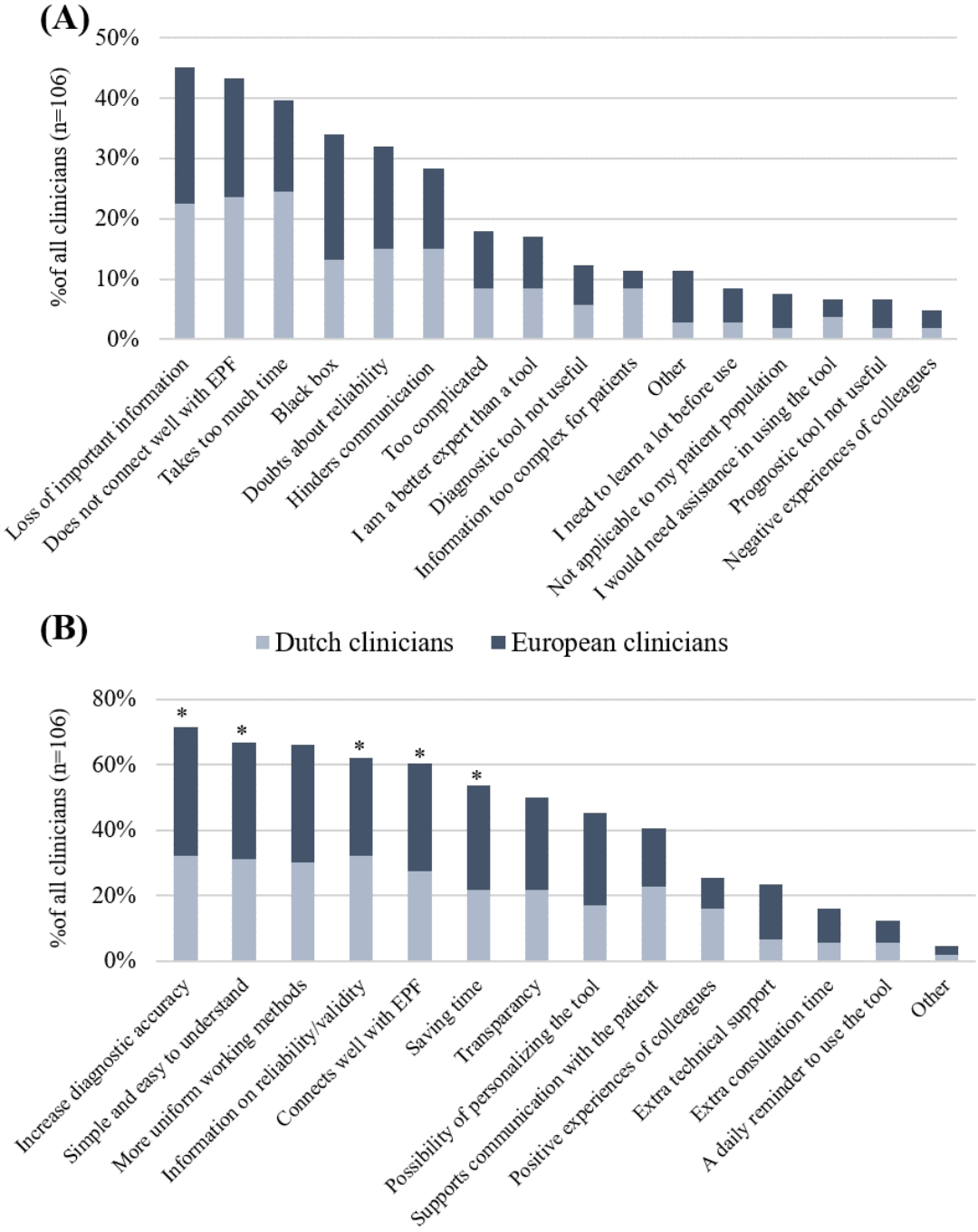
### *Barriers and Facilitators*

Clinicians' selection of factors that would hinder the use of computer tools has been displayed in Figure 1A. None of the barriers were rated using a mean Likert scale score ≥4. Figure 1B shows which factors would stimulate the use of computer tools; items marked with an asterisk were rated with a mean Likert scale score of ≥4. The most important barriers were (1) a tool not being well connected with the electronic patient file (47/109, 43.2% of clinicians), with a mean Likert scale score of 3.8 (SD 1.0), and (2) important information, such as observations, cannot be added to a tool (49/109, 44.9% of clinicians), with a mean Likert scale score of 3.7 (SD 1.2). The most important facilitators were (1) user-friendliness (73/109, 66.9% of clinicians; mean 4.5, SD 0.7), and (2) increasing diagnostic accuracy (78/109, 71.6% of clinicians; mean 4.3, SD 0.7).



XSL•FO
RenderX

**Figure 1.** Frequencies of barriers to (A) and facilitators of (B) the use of computer tools in daily practice according to Dutch and European clinicians. EPF: electronic patient file. *Items rated with a mean Likert scale score of ≥4.




### *Additional Tools for Memory Clinics*

Many clinicians (48/109, 44%) reported a willingness to use both web-based communication tools and skills training. Over half of the clinicians (62/109, 56.9%) indicated that they would like to test the patient's cognition via the computer at home before the appointment. Frequently selected reasons for web-based cognitive testing were triage (62/109, 56.9%) and shortening of the test battery at the clinic (37/109, 33.4%). Not being able to observe the patient during testing was mentioned by 89% (42/47) of the clinicians who would not want to test the patient's cognition on the web.

### *Interactive Panel Session*

Both transparencies about the objectives of the tool provider (Likert scale score mean 4.3, SD 1.1) and honesty about the possibilities and limitations of a tool (Likert scale scores mean 4.4, SD 0.9) were considered important to strengthen clinicians' trust in a tool. The most important aspects of convincing clinicians of usability, reliability, and validity of a tool were the explicit provision of information regarding the tool (mean 4.3, SD 0.9), scientific publications on underlying models (mean 4.3, SD 0.9), obtaining hands-on experience with the tool (mean 4.2, SD 0.6), and a randomized controlled trial (RCT) to test the effectiveness of the tool in clinical practice (mean 4.0, SD 0.8). Diagnostic accuracy (31/100, 31%) and patient-related



XSL•FO
RenderX

outcome measures (38/100, 38%), such as quality of life, were most frequently selected as the ideal primary outcome measures of such an RCT.

## Survey of Patients and Care Partners

### *Tools Used by Clinicians*

The results of patients' and care partners' opinions on diagnostic, prognostic, and communication tools have been presented in Table 4. Most patients and care partners were positive regarding their clinician using these tools. No differences were found between patients and care partners (diagnostic tools, *P*=.36; prognostic tools, *P*=.36; communication tools, *P*=.63) or different syndrome diagnoses (diagnostic tools, *P*=.64; prognostic tools, *P*=.69; communication tools, *P*=.92).

Figure 2 shows an overview of the responses of patients and care partners to several statements regarding the use of computer tools by clinicians. Items marked with an asterisk are items rated with a mean Likert scale score ≥4. We found no differences between the responses of patients and those of the care partners (*P* values in order of appearance of topics from top to down in the figure: *P*=.75, *P*=.62, *P*=.70, *P*=.55, *P*=.78, *P*=.21, and *P*=.60).

**Table 4.** Opinion of patients and care partners on clinicians' use of diagnostic, prognostic, and communication tools, illustrated with quotes (N=96).

| Opinion | Patients (n=50), n (%) | Care partners (n=46), n (%) | Quotes |
|---|---|---|---|
| **Diagnostic tool** | | | |
| I think that is a good thing | 38 (76) | 31 (67) | "The more information, the better. As long as the computer program is in addition to the doctor's expertise and not a replacement, I think it would be a good idea." [Female 60 years, care partner] |
| I would not want that | 4 (8) | 4 (9) | "I think face to face contact between the doctor and the patient is essential." [Female 74 years, care partner] |
| I do not know or no opinion | 8 (16) | 11 (24) | "Depends on how good the program is." [Male 76 years, patient, dementia] |
| **Prognostic tool** | | | |
| I think that is a good thing | 41 (82) | 32 (70) | "There is nothing against the use of a computer in predicting the disease process. It remains an aid to the physician. [...] He/she should remain leading." [Male 78 years, patient, SCD[a]] |
| I would not want that | 4 (8) | 6 (13) | "I want to know so I can plan ahead. However, with the variation in the progression rate, I don't see how this could be sufficiently accurate. If not accurate, I would not want it." [Female, 61 years, patient, dementia] |
| I do not know or no opinion | 5 (10) | 8 (17) | "My husband approves [the use of a prognostic tool], me as his wife, do not know if I would like it. What if the prediction is somber! We would instantly be depressed." [Female (age unknown), care partner] |
| **Communication tool** | | | |
| I think that is a good thing | 39 (78) | 38 (83) | —[b] |
| I would not want that | 3 (6) | 1 (2) | — |
| I do not know or no opinion | 8 (16) | 7 (15) | — |

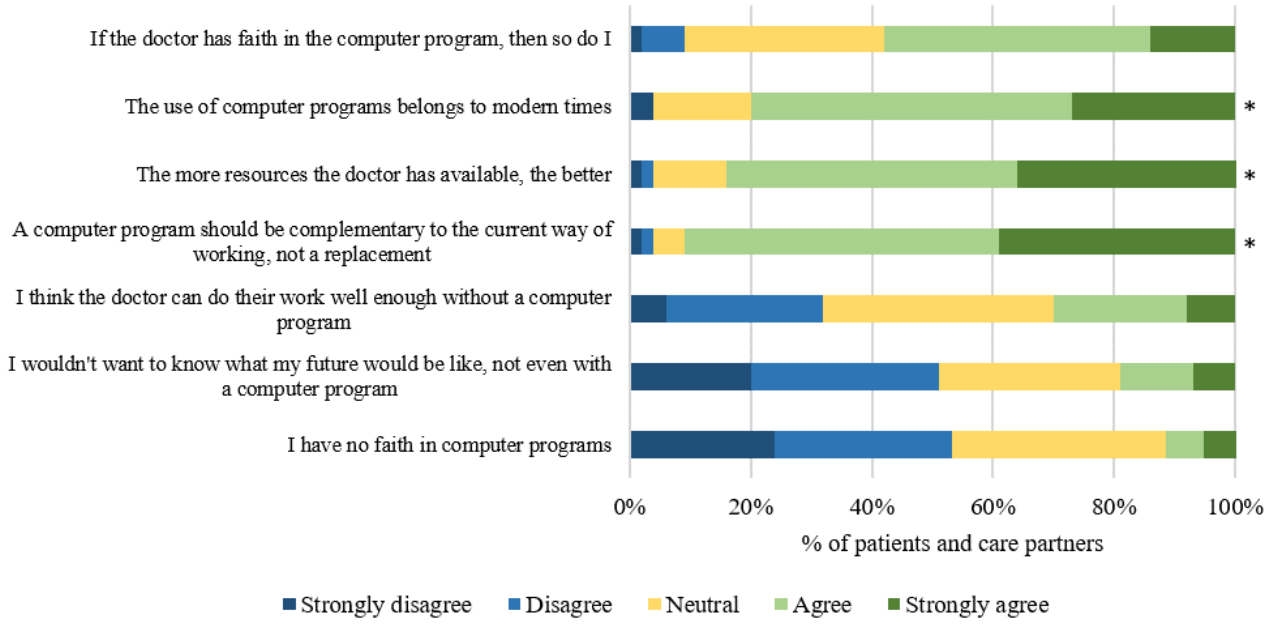
[a]SCD: subjective cognitive decline.

[b]No quotes available.



XSL•FO
RenderX

**Figure 2.** Agreement of patients and care partners on several statements regarding computer tools on a 5-point Likert scale. * Items rated with a mean Likert scale score of ≥4.




### *Tools Directed at Patients and Care Partners*

Most patients (35/50, 70%) and care partners (26/46, 57%) expressed a preference for a list of example questions to select questions they wanted to ask their clinician. Furthermore, 42% (21/50) of patients and 50% (23/46) of care partners said they would prepare their visit to the memory clinic by watching informational videos. A smaller but considerable proportion of care partners (11/46, 24%) indicated that they would use a web-based communication tool to practice their communication skills, and some (5/50, 10%) patients reported wanting to do the same. A small proportion of patients (4/50, 8%) and care partners (8/46, 17%) did not want to use any web-based communication tools.

Most patients (35/50, 70%) and care partners (28/46, 61%) were positive regarding web-based cognitive testing at home. Reasons for not wanting to perform cognitive testing at home differed between patients and care partners ($P$=.01). The most frequently selected reason for patients was their preference for personal contact with the clinician (18/26; 69%). The most selected reason (5/12, 42%) for care partners was that web-based cognitive testing was too difficult to perform for their loved ones.

## Discussion

### Principal Findings

This study showed that most clinicians, patients, and care partners were supportive of the use of computer tools in memory clinics. This holds true for diagnostic and prognostic tools, tools that support communication, and web-based cognitive testing. Despite acknowledging their potential barriers, the general attitude of clinicians toward these tools was positive. The facilitating factors were mainly practical (tools should be user-friendly) and technical (connection with electronic patient files) and that tools should increase diagnostic accuracy. The identified barriers mainly focused on doubts regarding reliability and validity, preservation of clinical autonomy, and fear of losing important clinical information. Furthermore, the tools should be considered in addition to the current working methods and not as a replacement.

We hypothesized that the barriers to and facilitators of tools in memory clinics might differ from those identified in other health care areas because of the nature of the patient population (older adults with cognitive decline) and the large number of diagnostic tests used in the clinical workup of dementia. However, the barriers and facilitators we found in this survey study largely corresponded to the existing literature on barriers to and facilitators of the use of computer tools in other medical fields [27-33]. Furthermore, we found that most of the given answers to the open-ended question were in accordance with our predefined list, which was based on studies on applying computer tools in other health care areas [27,28,32]. In addition, it is conceivable that patients' and their care partners' (negative) opinions regarding the tools were a possible barrier to clinicians using a tool. In this study, we showed that patients, despite their age and (potential) cognitive decline, are mainly positive regarding the use of computer tools. Most of them embrace the possibility of using a tool themselves, and their care partners share this opinion. A computer tool must support rather than replace clinicians, who would then view it as an aid appropriate to modern times.

### Increase Acceptance of Tools

We found the overall attitude among clinicians toward tools to be highly positive. Nevertheless, none of the available tools are regularly used in daily practice, and it seems there is a major information gap and educational need to make clinicians understand the possibilities of such computer tools [43]. The results from our survey provide direction for the way to increase



XSL•FO
RenderX

the acceptance of computer tools in memory clinic practice. First, clinicians indicated that an RCT on the efficiency in clinical practice would boost their confidence in the reliability of computer tools. RCTs are considered important and robust methods for assessing the impact of a tool [44,45]. However, no RCT on the application of computer tools has yet been performed in memory clinics. Thus, our findings suggest a need for RCTs with diagnostic accuracy and patient-related outcome measures as primary outcome measures.

Second, computer tools could contribute to evidence-based medicine (EBM) [42]. EBM concerns medical practice based on the best available evidence, clinical experience, and patient preferences [42]. Within the concept of EBM, clinical experience is highly essential and should not be replaced by a tool. Computer tools could support EBM by making the best available evidence more accessible to clinicians, and computer tools could clarify patient preferences. Acknowledging computer tools as a part of EBM might lead to clinicians viewing these tools as an aid complementary to their own clinical experience rather than a threat to their clinical autonomy [27,29,32,46]. In addition, there must be scientific publications regarding the underlying models and the transparent provision of information regarding a tool's reliability to increase clinicians' trust in the tools [28,32,47].

Then, clinicians' confidence in the tools might be strengthened if the tools are under the jurisdiction of a regulatory body to authorize and supervise the quality. To date, there are no formal regulatory standards for tools used to support clinicians in decision-making [28]. When developers claim that their software has a medical purpose, it becomes a medical device, and manufacturers themselves have to proclaim that their device meets the safety and performance requirements. The new European Medical Devices Regulation implemented in May 2021 makes manufacturers adhere to more strict guidelines for ensuring the safety of their products, including assessment of their device by a notified body. Involving an independent notified body in the approval of a tool might be the first step toward improving its transparency and acceptance. Eventually, as a next step, the use of software should be included in the guidelines of professional associations [26].

The attitude toward tools is one of the key characteristics of eventual acceptance [30,32]. Several user acceptance models have been proposed to further encourage the acceptance of tools in medical practices [32]. In this context, we would like to address the user acceptance and system adaptation design model [32]. This model aims to include end users as the central point in the design process of a computer tool. In this model, user expectations and needs need to be thoroughly understood before developing a tool. The development of a tool is an iterative process, and end users should be continuously involved throughout the development process. In congruence with this user acceptance model, we took the first step toward accepting computer tools by identifying the barriers to and facilitators of computer tools according to the end users in memory clinics. On the basis of the user acceptance and system adaptation design model, the next step in the iterative process would be the evaluation of usability. Then, a pilot study would be needed to evaluate user acceptance. The results of each step should direct improvements in the tool and be evaluated in the next step. However, it should be mentioned that in this study, we asked participants about the barriers to and facilitators of computer tools in general. Therefore, the barriers and facilitators found in this study should be considered a starting point. When implementing specific tools, further exploration of user expectations and needs for those specific tools might be necessary.

## Strengths and Limitations

One of the strengths of this study is that we involved all the end users of computer tools in memory clinics, that is, clinicians, patients, and care partners. Including patients and care partners originating from Europe and a heterogeneous population of clinicians contributed to the generalizability of the results. Furthermore, we used a funneled method in which we started the survey with open-ended questions and worked toward closed questions. By doing so, we actively asked and stimulated clinicians' own input.

Nevertheless, our study has several limitations. First, we distributed the survey via a web-based link, which might have caused selection bias by only involving people with sufficient digital skills who might have had a more positive approach toward computer tools. We tried to minimalize this risk for patients and care partners by including participants from both geriatric and neurology departments with different cognitive impairment stages. Furthermore, participants' ages ranged from relatively young to older patients, who might have had less experience with digital tools. Second, we have no data available on the origin of the European participants, which might have led to an uneven distribution of participants among countries. Nonetheless, the international character of this study contributes to the generalizability of the results. Third, there might be a risk of response bias; people who are less inclined to use computer tools may not have responded. We could not estimate the response rates as the link to the survey was spread among an unknown number of people. Nonetheless, we gained insight into the important barriers and facilitators based on inquiry among large samples of the most important stakeholders.

## Conclusions

In conclusion, this study shows broad support for the use of computer tools in memory clinic practices by clinicians, patients, and care partners. To stimulate the implementation of tools in daily memory practice, the tools should overcome several technical and practical barriers. Moreover, clinicians have to be convinced regarding the reliability and validity of the tool. By identifying the potential barriers and facilitators, we have paved the way for further development and implementation of the tools. Our results provide an important step in the iterative process of developing computer tools for memory clinics in cocreation with end users.



XSL•FO
RenderX

## Acknowledgments

Research of the Alzheimer's Center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. The Vrije Universiteit Medical Center Alzheimer Center is supported by the Stichting Alzheimer Nederland and Stichting Vrije Universiteit Medical Center Fonds. The clinical database structure was developed with funding from the Stichting Dioraphte. The collaboration project Dementia diagnostics using Artificial Intelligence (light weighted computer tool for local memory clinics; project number LSHM19123-HSGF) is cofunded by the public–private partnership allowance made available by Health-Holland, Top Sector Life Sciences and Health, to stimulate public–private partnerships. The chair of WMVDF is supported by the Pasman Stichting. WMVDF and LNCV are recipients of ABOARD (A personalized medicine approach for Alzheimer disease), which is a public–private partnership receiving funding from ZonMW (number 73305095007) and Health-Holland, Top Sector Life Sciences and Health (public–private partnership allowance; number LSHM20106). LNCV is supported by a fellowship grant received from Alzheimer Nederland (WE.15-2019-05). WF, LNV and JG are recipients of the EU Joint Programme - Neurodegenerative Disease Research (JPND) project EURO-FINGERS (ZonMW-Memorabel #733051102) which is supported through the following funding organizations under the aegis of Joint Programme-Neurodegenerative Disease: Finland, Academy of Finland; Germany, Federal Ministry of Education and Research; Spain, National Institute of Health Carlos III; Luxemburg, National Research Fund; Hungary, National Research, Development and Innovation Office; The Netherlands, Netherlands Organisation for Health Research and Development; Sweden, Swedish Research Council (grant agreement INTER/JPND/19/BM/14012609). These funding sources were not involved in the study design, collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

## Conflicts of Interest

HFMRM performs contract research for Combinostics; all funding is paid to her institution. WMVDF performs contract research for Biogen. Research programs of WMVDF have been funded by ZorgOnderzoek Nederland, area of Medical Sciences, Dutch Research Council, European Union Framework Programmes, Alzheimer Nederland, Cardiovascular Onderzoek Nederland, Gieskes-Strijbis fonds, Pasman Stichting, Boehringer Ingelheim, Piramal Neuroimaging, Combinostics, Roche BV, AVID. She has been an invited speaker at Boehringer Ingelheim and Biogen. All funding is paid to her institution. FHB performs contract research for Optina Dx and Optos, she has been an invited speaker at Roche and has been invited for expert testimony at Biogen. All funding is paid to her institution. All other coauthors report no conflicts of interest.

## References

1. The epidemiology and impact of dementia: current state and future trends. World Health Organization. 2015. URL: https://www.who.int/mental_health/neurology/dementia/dementia_thematicbrief_epidemiology.pdf [accessed 2020-11-16]
2. Dubois B, Padovani A, Scheltens P, Rossi A, Dell'Agnello G. Timely diagnosis for Alzheimer's disease: a literature review on benefits and challenges. J Alzheimers Dis 2016;49(3):617-631 [FREE Full text] [doi: 10.3233/JAD-150692] [Medline: 26484931]
3. Prince M, Ferri C. World Alzheimer Report 2011: the benefits of early diagnosis and intervention. Alzheimer's Disease International (ADI) and Institute of Psychiatry, King's College London. 2011. URL: https://www.alzint.org/u/WorldAlzheimerReport2011.pdf [accessed 2021-10-29]
4. Robinson L, Tang E, Taylor J. Dementia: timely diagnosis and early intervention. Br Med J 2015 Jun 16;350:h3029 [FREE Full text] [doi: 10.1136/bmj.h3029] [Medline: 26079686]
5. van Maurik IS, Vos SJ, Bos I, Bouwman FH, Teunissen CE, Scheltens P, et al. Biomarker-based prognosis for people with mild cognitive impairment (ABIDE): a modelling study. Lancet Neurol 2019 Nov;18(11):1034-1044. [doi: 10.1016/s1474-4422(19)30283-2]
6. Scheltens P, De Strooper B, Kivipelto M, Holstege H, Chételat G, Teunissen CE, et al. Alzheimer's disease. Lancet 2021 Apr;397(10284):1577-1590. [doi: 10.1016/s0140-6736(20)32205-4]
7. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):263-269 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]
8. Román GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology 1993 Mar;43(2):250-260. [doi: 10.1212/wnl.43.2.250] [Medline: 8094895]
9. Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 2011 Sep;134(Pt 9):2456-2477 [FREE Full text] [doi: 10.1093/brain/awr179] [Medline: 21810890]
10. McKeith IG, Boeve BF, Dickson DW, Halliday G, Taylor J, Weintraub D, et al. Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. Neurology 2017 Jul 04;89(1):88-100 [FREE Full text] [doi: 10.1212/WNL.0000000000004058] [Medline: 28592453]



XSL•FO
RenderX

11. Vos SJ, Verhey F, Frölich L, Kornhuber J, Wiltfang J, Maier W, Alzheimer's Disease Neuroimaging Initiative, et al. Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage. Brain 2015 May 17;138(Pt 5):1327-1338 [FREE Full text] [doi: 10.1093/brain/awv029] [Medline: 25693589]
12. Fruijtier AD, Visser LN, Bouwman FH, Lutz R, Schoonenboom N, Kalisvaart K, et al. What patients want to know, and what we actually tell them: the ABIDE project. Alzheimers Dement (N Y) 2020;6(1):e12113 [FREE Full text] [doi: 10.1002/trc2.12113] [Medline: 33344753]
13. Visser LN, Pelt SA, Kunneman M, Bouwman FH, Claus JJ, Kalisvaart KJ, et al. Communicating uncertainties when disclosing diagnostic test results for (Alzheimer's) dementia in the memory clinic: the ABIDE project. Health Expect 2020 Feb;23(1):52-62 [FREE Full text] [doi: 10.1111/hex.12964] [Medline: 31638322]
14. Kunneman M, Pel-Littel R, Bouwman FH, Gillissen F, Schoonenboom NS, Claus JJ, et al. Patients' and caregivers' views on conversations and shared decision making in diagnostic testing for Alzheimer's disease: the ABIDE project. Alzheimers Dement (N Y) 2017 Sep;3(3):314-322 [FREE Full text] [doi: 10.1016/j.trci.2017.04.002] [Medline: 29067338]
15. Visser LN, Kunneman M, Murugesu L, van Maurik I, Zwan M, Bouwman FH, et al. Clinician-patient communication during the diagnostic workup: the ABIDE project. Alzheimers Dement (Amst) 2019 Dec;11:520-528 [FREE Full text] [doi: 10.1016/j.dadm.2019.06.001] [Medline: 31388556]
16. Bruun M, Frederiksen KS, Rhodius-Meester HF, Baroni M, Gjerum L, Koikkalainen J, et al. Impact of a clinical decision support tool on dementia diagnostics in memory clinics: the PredictND validation study. Curr Alzheimer Res 2019;16(2):91-101. [doi: 10.2174/1567205016666190103152425] [Medline: 30605060]
17. van Maurik IS, Visser LN, Pel-Littel RE, van Buchem MM, Zwan MD, Kunneman M, et al. Development and usability of adappt: web-based tool to support clinicians, patients, and caregivers in the diagnosis of mild cognitive impairment and Alzheimer disease. JMIR Form Res 2019 Jul 08;3(3):e13417 [FREE Full text] [doi: 10.2196/13417] [Medline: 31287061]
18. Tolonen A, Rhodius-Meester HF, Bruun M, Koikkalainen J, Barkhof F, Lemstra AW, et al. Data-driven differential diagnosis of dementia using multiclass disease state index classifier. Front Aging Neurosci 2018 Apr 25;10:111 [FREE Full text] [doi: 10.3389/fnagi.2018.00111] [Medline: 29922145]
19. Rhodius-Meester HF, van Maurik IS, Koikkalainen J, Tolonen A, Frederiksen KS, Hasselbalch SG, et al. Selection of memory clinic patients for CSF biomarker assessment can be restricted to a quarter of cases by using computerized decision support, without compromising diagnostic accuracy. PLoS One 2020;15(1):e0226784 [FREE Full text] [doi: 10.1371/journal.pone.0226784] [Medline: 31940390]
20. Zygouris S, Tsolaki M. Computerized cognitive testing for older adults: a review. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):13-28. [doi: 10.1177/1533317514522852] [Medline: 24526761]
21. Rhodius-Meester HF, Paajanen T, Koikkalainen J, Mahdiani S, Bruun M, Baroni M, et al. cCOG: a web-based cognitive test tool for detecting neurodegenerative disorders. Alzheimers Dement (Amst) 2020;12(1):e12083 [FREE Full text] [doi: 10.1002/dad2.12083] [Medline: 32864411]
22. Hansen TI, Haferstrom EC, Brunner JF, Lehn H, Håberg AK. Initial validation of a web-based self-administered neuropsychological test battery for older adults and seniors. J Clin Exp Neuropsychol 2015;37(6):581-594 [FREE Full text] [doi: 10.1080/13803395.2015.1038220] [Medline: 26009791]
23. Licqurish SM, Cook OY, Pattuwage LP, Saunders C, Jefford M, Koczwara B, et al. Tools to facilitate communication during physician-patient consultations in cancer care: an overview of systematic reviews. CA Cancer J Clin 2019 Nov;69(6):497-520 [FREE Full text] [doi: 10.3322/caac.21573] [Medline: 31339560]
24. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Dec 12;4:CD001431. [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
25. Frisoni GB, Boccardi M, Barkhof F, Blennow K, Cappa S, Chiotis K, et al. Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. Lancet Neurol 2017 Dec;16(8):661-676. [doi: 10.1016/S1474-4422(17)30159-X] [Medline: 28721928]
26. Turner RS, Stubbs T, Davies DA, Albensi BC. Potential New Approaches for Diagnosis of Alzheimer's Disease and Related Dementias. Front Neurol 2020 Jun 5;11:496 [FREE Full text] [doi: 10.3389/fneur.2020.00496] [Medline: 32582013]
27. Kilsdonk E, Peute LW, Jaspers MW. Factors influencing implementation success of guideline-based clinical decision support systems: a systematic review and gaps analysis. Int J Med Inform 2017 Dec;98:56-64. [doi: 10.1016/j.ijmedinf.2016.12.001] [Medline: 28034413]
28. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. J Am Med Assoc 2018 Dec 04;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
29. Varonen H, Kortteisto T, Kaila M, EBMeDS Study Group. What may help or hinder the implementation of computerized decision support systems (CDSSs): a focus group study with physicians. Fam Pract 2008 Jun;25(3):162-167. [doi: 10.1093/fampra/cmn020] [Medline: 18504253]
30. Devaraj S, Sharma SK, Fausto DJ, Viernes S, Kharrazi H. Barriers and facilitators to clinical decision support systems adoption: a systematic review. J Busi Admin Res 2014 Jul 24;3(2):5217. [doi: 10.5430/jbar.v3n2p36]




XSL•FO
RenderX

31. Kortteisto T, Komulainen J, Mäkelä M, Kunnamo I, Kaila M. Clinical decision support must be useful, functional is not enough: a qualitative study of computer-based clinical decision support in primary care. BMC Health Serv Res 2012 Oct 08;12:349 [FREE Full text] [doi: 10.1186/1472-6963-12-349] [Medline: 23039113]
32. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
33. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
34. Survalyzer - to survey and analyze. Utrecht. URL: https://education.survalyzer.com/ [accessed 2020-09-11]
35. Mentimeter. Mentimeter AB, Stockholm, Sweden. URL: https://mentimeter.com [accessed 2020-12-14]
36. van der Flier WM, Pijnenburg YA, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. J Alzheimers Dis 2014;41(1):313-327. [doi: 10.3233/JAD-132306] [Medline: 24614907]
37. van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: performing research to optimize care. J Alzheimers Dis 2018;62(3):1091-1111 [FREE Full text] [doi: 10.3233/JAD-170850] [Medline: 29562540]
38. Rhodius-Meester HF, van de Schraaf SA, Peters MJ, Kleipool EE, Trappenburg MC, Muller M. Mortality risk and its association with geriatric domain deficits in older outpatients: the Amsterdam Ageing Cohort. Gerontology 2021 Jan 13;67(2):194-201 [FREE Full text] [doi: 10.1159/000512048] [Medline: 33440389]
39. MAXQDA Standard 18.2.0 ed. URL: https://www.maxqda.com/new-maxqda-2018 [accessed 2021-10-29]
40. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2016 Dec 09;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
41. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
42. Masic I, Miokovic M, Muhamedagic B. Evidence based medicine - new approaches and challenges. Acta Inform Med 2008;16(4):219-225 [FREE Full text] [doi: 10.5455/aim.2008.16.219-225] [Medline: 24109156]
43. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008 Apr;41(2):387-392 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.003] [Medline: 18029232]
44. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 3;157(1):29-43. [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
45. Liu JL, Wyatt JC. The case for randomized controlled trials to assess the impact of clinical information systems. J Am Med Inform Assoc 2011;18(2):173-180 [FREE Full text] [doi: 10.1136/jamia.2010.010306] [Medline: 21270132]
46. Murray-Torres T, Casarella A, Bollini M, Wallace F, Avidan MS, Politi MC. Anesthesiology Control Tower-Feasibility Assessment to Support Translation (ACTFAST): mixed-methods study of a novel telemedicine-based support system for the operating room. JMIR Hum Factors 2019 Apr 23;6(2):e12155 [FREE Full text] [doi: 10.2196/12155] [Medline: 31012859]
47. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]

## Abbreviations

**EBM:** evidence-based medicine
**NGN:** Nederlands Geheugenpoli Netwerk
**RCT:** randomized controlled trial
**SCD:** subjective cognitive decline

*Edited by G Eysenbach; submitted 08.06.21; peer-reviewed by S Hasselbalch, D Hu; comments to author 08.08.21; revised version received 07.09.21; accepted 18.09.21; published 03.12.21.*

*Please cite as:*
*van Gils AM, Visser LNC, Hendriksen HMA, Georges J, Muller M, Bouwman FH, van der Flier WM, Rhodius-Meester HFM Assessing the Views of Professionals, Patients, and Care Partners Concerning the Use of Computer Tools in Memory Clinics: International Survey Study*
*JMIR Form Res 2021;5(12):e31053*
*URL: https://formative.jmir.org/2021/12/e31053*
*doi:10.2196/31053*
*PMID:34870612*



XSL•FO
RenderX

©Aniek M van Gils, Leonie NC Visser, Heleen MA Hendriksen, Jean Georges, Majon Muller, Femke H Bouwman, Wiesje M van der Flier, Hanneke FM Rhodius-Meester. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# A Multimodal Messaging App (MAAN) for Adults With Autism Spectrum Disorder: Mixed Methods Evaluation Study

Mohamad Hassan Fadi Hijab[1], MSc; Dena Al-Thani[1], PhD; Bilikis Banire[1], PhD

Division of Information and Computer Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

**Corresponding Author:**
Mohamad Hassan Fadi Hijab, MSc
Division of Information and Computer Technology
College of Science and Engineering
Hamad Bin Khalifa University
Qatar Foundation
Education City
Doha, 34110
Qatar
Phone: 974 50677321
Email: mhhijab@hbku.edu.qa

## Abstract

**Background:** Individuals with autism spectrum disorder (ASD) often exhibit difficulties in social and communication skills. For more than 30 years, specialists, parents, and caregivers have used techniques, such as applied behavioral analysis, augmentative and alternative communication, and the picture exchange communication system to support the social and communication skills of people with ASD. Even though there are many techniques devised to enhance communication, these techniques are not considered in existing social media apps for people with ASD.

**Objective:** This study aimed to investigate the effect of adding accessibility features, such as text-to-speech (TTS), speech-to-text (STT), and communication symbols (CS), to a messaging app (MAAN). We hypothesized that these accessibility features can enhance the social and communication skills of adults with ASD. We also hypothesized that usage of this app can reduce social loneliness in adults with ASD.

**Methods:** Semistructured interviews were conducted with 5 experts working in fields related to ASD to help design the app. Seven adults with ASD participated in the study for a period of 10 to 16 weeks. Data logs of participants' interactions with the app were collected. Additionally, 6 participants' parents and 1 caregiver were asked to complete a short version of the Social and Emotional Loneliness Scale for Adults (SELSA-S) questionnaire to compare pre-post study results. The Mobile Application Rating Scale: user version questionnaire was also used to evaluate the app's usability. Following the study, interviews were conducted with participants to discuss their experiences with the app.

**Results:** The SELSA-S questionnaire results showed no change in the family subscale; however, the social loneliness subscale showed a difference between prestudy and poststudy. The Wilcoxon signed-rank test indicated that poststudy SELSA-S results were statistically significantly higher than prestudy results ($z=-2.047$; $P=.04$). Point-biserial correlation indicated that the SELSA-S rate of change was strongly related to usage of the TTS feature ($r=0.708$; $P=.04$) and CS feature ($r=-0.917$; $P=.002$), and moderately related to usage of the STT feature ($r=0.428$; $P=.17$). Lastly, we adopted grounded theory to analyze the interview data, and the following 5 categories emerged: app support, feature relevance, user interface design, overall feedback, and recommendations.

**Conclusions:** This study discusses the potential for improving the communication skills of adults with ASD through special features in mobile messaging apps. The developed app aims to support the inclusion and independent life of adults with ASD. The study results showed the importance of using TTS, STT, and CS features to enhance social and communication skills, as well as reduce social loneliness in adults with ASD.

***(JMIR Form Res 2021;5(12):e33123)*** doi:10.2196/33123

**KEYWORDS**

autism; assistive technology; mobile app; social and communication skills



XSL•FO
RenderX

## Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that is characterized by repetitive patterns of behavior, restricted interest, and deficits in social and communication skills [1]. These deficits affect the mental development of people with ASD, especially their socioemotional development, which may result in social and family loneliness [2]. Loneliness can be described as a lack of a close or intimate attachment to family members and friends. The consequences of loneliness include feelings of emptiness, anxiety, depression, suicide, and anger, which are detrimental to both physical and mental health [3]. Another effect of loneliness is learning disabilities, as the deficits associated with ASD limit patients' opportunities to interact with their social environment and result in delays in learning social norms [3]. Understanding social norms is essential for positive and long-lasting social relationships. For example, an individual with ASD might not know when someone is angry and how to reciprocate. This lack of social norms may lead to social rejection. Social and communication skills are thus crucial components for social and emotional development. In recent years, the ASD prevalence rate has increased, with significant variability worldwide. In the United States, 185 in 10,000 children have a diagnosis of ASD [4]. Similarly, recent prevalence studies in the Middle East reported that 114 in 10,000 children from Qatar [5], and between 49 and 513 in 10,000 children from Lebanon have the disease [6]. Given the increasing prevalence of ASD, the impact of problems in social and communication skills associated with this disorder may lead to a high societal cost of supporting the physical and mental development of individuals with ASD. Thus, it is critical to develop innovative communication platforms to enhance social and communication skills between individuals with ASD and others.

Many interventions have been used to enhance social and communication skills, such as applied behavioral analysis (ABA) [7], augmentative and alternative communication (AAC) [8], and the picture exchange communication system (PECS) [9]. ABA is known as the primary treatment for children with ASD [10]. It is considered necessary for education and behavioral interventions, as it is commonly used to assist and improve social and communication skills for children with ASD. AAC provides individuals with ASD with pictures and illustrations to express their needs, and it is extensively used, although some caregivers and specialists have expressed concerns regarding its application. They often highlight that usage of AAC will prohibit the development of verbal speech in individuals with ASD [11]. Nevertheless, these technologies have the potential for supporting the social and communication skills of individuals with ASD [12]. The number of studies using AAC has increased due to the rise in its demand and value for an assistive technology tool in the last few years [13]. Lastly, the PECS is considered as one of the forms of AAC to help individuals with ASD to communicate. These systems appear promising due to their ability to construct sentences using images and symbols [14], with a low requirement of minor motor planning skills and little cognitive demand, which is why limited training is needed [15]. Many studies have investigated the importance of the PECS and determined its major role in enhancing social and communication skills [14,16]. On the other hand, PECS has several disadvantages, including difficulties in virtual communication and low speed in constructing messages [15]. However, behavioral and educational interventions, and digital revolutions have tremendously influenced the lives of individuals with ASD positively [17]. These technologies have increased the utilization of virtual communication platforms and highlighted the importance of providing remote services to the population with ASD [18]. Moreover, technology can ease communication and enhance individuals' social well-being [19]. The use of assistive technologies and tools can help create a platform that allows individuals with ASD to share their thoughts, ideas, and emotions effortlessly [19]. These technologies were introduced to support their communication challenges, and help them interact and express their feelings with the world without physical presence [20].

This study intends to verify the effectiveness of a new mobile messaging app called "MAAN," which means "together" in Arabic, for adults with ASD, using pre-post study data. The data were collected using questionnaires, interviews, and interaction logs. The aim of MAAN is to provide an accessible, safe, and easy-to-use messaging app for adults with ASD and hence support their inclusion in society. In this study, we hypothesized the following: (1) The accessibility features of mobile apps (text-to-speech [TTS], speech-to-text [STT], and communication symbols [CS]), which are applied to MAAN, can enhance the social and communication skills of adults with ASD; and (2) The usage of the app can reduce social loneliness in adults with ASD.

## Methods

### Study Design

A mixed methods approach was followed in this study. Seven adults with ASD were asked to use MAAN with their caregivers for 16 weeks. Prior to the design and development of MAAN, semistructured interviews were conducted with experts to discuss the preliminary proposed app design. The app design was then modified according to the received feedback. After that, adults with ASD and their parents or caregivers were asked to use MAAN for a period of 16 weeks. This duration was adopted from the study by Laugeson et al [21], which suggested that when receiving social and communication skills interventions, improvement is only evident at 16 weeks. The study started by contacting individuals willing to participate. Due to COVID-19, the initial meetings with the participants were conducted online. All participants received an email for explaining the study information, such as date and time, and a weblink for collecting informed consent for a virtual meeting with the participants, their friends, and their parents. After receiving the participants' consent, they were asked to fill an online prestudy questionnaire in order to collect demographical information. The parents or caregivers of the participants were also asked to complete a short version of the Social and Emotional Loneliness Scale for Adults (SELSA-S) questionnaire [21]. The aim of this questionnaire was to determine the social and family loneliness of the participants. In a virtual meeting,



XSL•FO
RenderX

the principal researcher explained the app features in detail and answered further questions regarding the study. The participants were also provided with a hotline to report any issue they might face during this study. To ensure the continuity of data collection, the participants were asked to make sure that the app was constantly running in the background.

At the end of the study, all participants with ASD and their caregivers or parents completed the Mobile Application Rating Scale: user version (uMARS) [22] questionnaire. The uMARS is a simple and reliable tool that can be used by end-users to assess the quality of mobile health apps. It provides a 20-item measure that includes 4 objective quality subscales (engagement, functionality, esthetics, and information quality) and 1 subjective quality subscale. The purpose of this questionnaire is to measure the app's quality in terms of usability. The participants' parents or caregivers also completed the SELSA-S questionnaire again to determine the participants' social and family loneliness rates after the study. Lastly, a semistructured interview with parents or caregivers was conducted to gather information about the context of app usage. The interview included questions about the overall app experience in terms of usability, functionality, and challenges encountered while using the app, as well as suggestions for further enhancement.

## MAAN App Design and Development

The app consisted of the following 3 main features: TTS, STT, and CS. One of the well-known features that can also be considered as a type of assistive technology is the TTS feature, which allows users to listen to written text on a computer, tablet, or smartphone. This technology is popular among people having difficulties in reading and decoding words [23]. Similarly, STT is an assistive technology feature used by people who struggle with writing. This form of assistive technology allows speech to be converted into text using the computing device. Both features are very commonly employed to support inclusion education settings [24]. The app also included the CS feature, which is an essential part of AAC-supported technologies, and its primary purpose is to assist individuals with ASD in constructing sentences. We are not aware of any messaging app that supports the use of the CS feature. "TAWASOL symbols" were employed for the CS feature. They were developed in 2013 by Mada Center, an assistive technology center in Qatar. These symbols represent the Modern Standard Arabic language and are designed to be culturally, socially, religiously, and linguistically acceptable. TAWASOL is based on the AAC Symbols Collection and is directed to nonverbal or minimally verbal individuals who require alternative communication solutions [25].

MAAN was developed using Xcode, which is an Apple integrated development environment (IDE) for developing iOS-based software. MAAN is available on iOS devices and is designed with a minimal number of screens. The screen designs are similar to traditional and commonly used messaging apps. Thus, in order to provide individuals with a consistent and effortless experience, TTS and STT features were implemented in the app using Xcode libraries. Additionally, 478 images from TAWASOL symbols were divided into 25 categories and added to MAAN. The essence of the category was to reduce the search time of required symbols. Participants can click on a certain category and then choose a TAWASOL symbol. The textual description of the selected TAWASOL symbol will be displayed in both Arabic and English. The user can specify the language of the symbol textual description, which is another advantage of MAAN over other apps. When previewing the received symbol, the user can listen to the TAWASOL symbol textual description. A challenging aspect related to designing for adults with ASD is the choice of the interface color theme [26,27]. MAAN aims to address this issue by allowing users to choose their suitable color theme. The MAAN app flow can be divided into the following 2 sections: administrative and features. Figure 1 shows screenshots of the app's administrative flow, and Figure 2 shows screenshots of the app's features.

**Figure 1.** Screenshots of the MAAN app's administrative section.

In the administrative section, the app starts with a login screen where the participants can login using an assigned username and password, and remain logged in until they logout manually (Figure 1A). The participants can select whom to chat with and

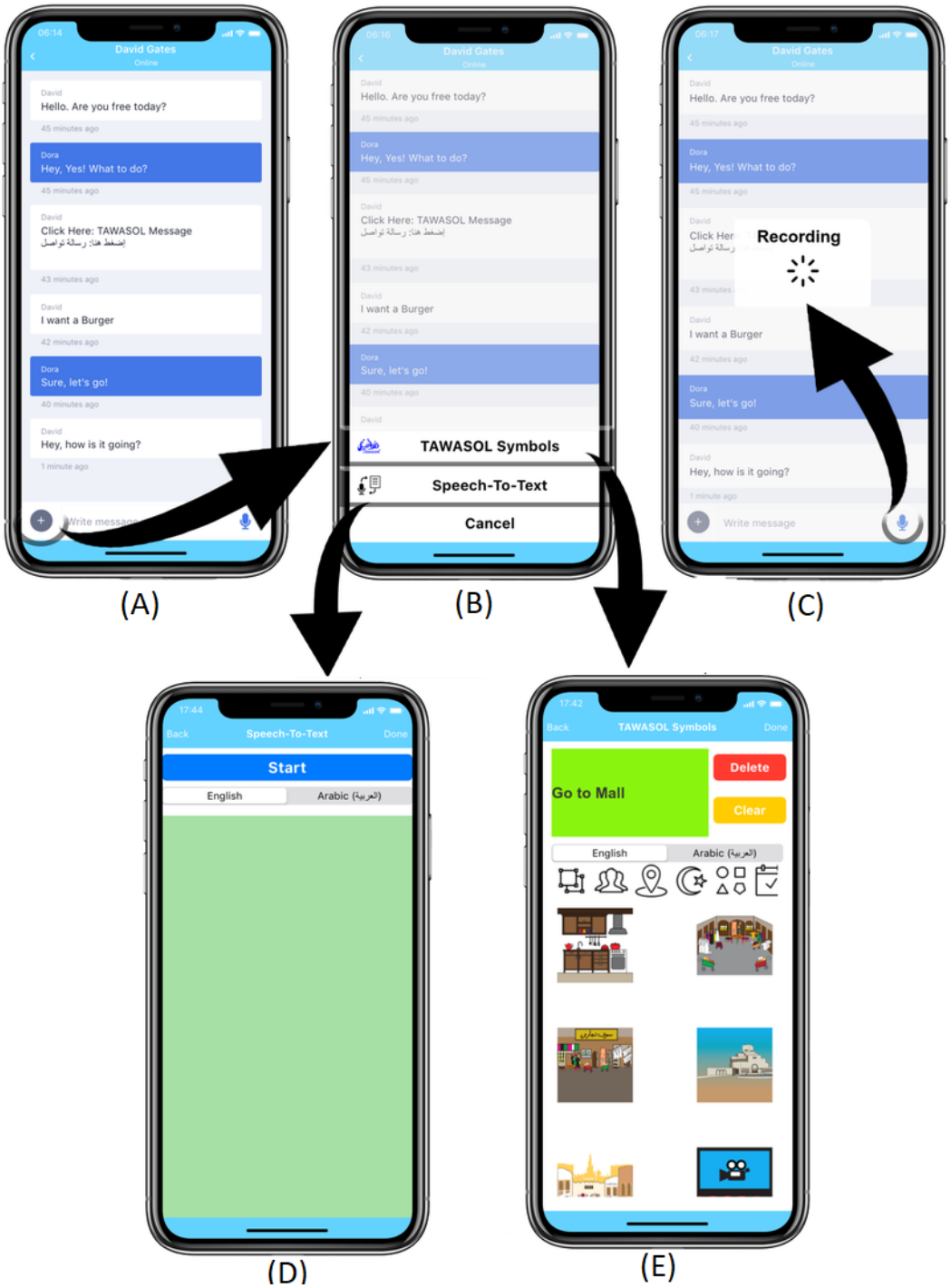


XSL•FO
RenderX

start their conversation (Figure 1B). The setting screen provides the participants with the ability to view their profile, and change the app theme color and the incoming and outgoing messaging bubble color (Figure 1C and Figure 1D). Moreover, the participants can change the language of the app (Arabic or English) and logout from the app (Figure 1C).

The participants can use the features of the app as described in Figure 2. An accessory button in the chat window provides the participants with the choice to select the STT or TAWASOL symbols screen (Figure 2A and Figure 2B). On clicking the mic icon, they can record their voice and send it as a voice note to the person they are chatting with (Figure 2C). On selecting STT, the participants proceed to another screen where they can chat, and the chat will be directly transcribed into text to be sent (Figure 2D). On selecting TAWASOL symbols, the participants can choose symbols from the 25 different categories, which will be generated as text to be sent (Figure 2E). The last feature, TTS, will be activated when the participants press on any text message. With this feature, the system will transcribe the text into voice and read it.

**Figure 2.** Screenshots of the MAAN app's feature display screens.






XSL•FO
RenderX

### Testing and Analysis Tools

Ethical approval was obtained from the Research Board of Qatar Biomedical Research Institute. Participation was entirely voluntary, and each participant was sent an approved informed consent form that contained all study details. The following 4 types of consent forms were sent based on participant category: (1) participant without ASD (including caregiver or parent) consent form, (2) specialist consent form, (3) verbal participant with ASD consent form, and (4) nonverbal participant with ASD consent form. Each participant was aware of the requirement to provide an informed decision regarding participation in the study, as well as the right to withdraw from the study without justification or penalties. Moreover, participants were assured about the confidentiality and security of the data collected. The user evaluation study was conducted in a real-life setting, and the participants were recruited through snowballing techniques [28]. As of November 2020, a total of 7 adults with ASD, aged between 18 and 30 years, were recruited from Qatar and Lebanon. For each participant, parents, caregivers, or specialists were recruited for using the app.

Quantitative and qualitative measures were collected in this research. In terms of quantitative measures, this study employed 3 questionnaires and collected the interaction log data through the app. The prestudy questionnaire was used to collect the participants' demographic data, and the SELSA-S [29,30] questionnaire was used to determine social and family loneliness for participants prior to the study and after completing the study. The uMARS [31] questionnaire was used to measure the app's quality after completing the study. Interaction log data were also collected through Google Firebase. This included the number of times the app was accessed, and how each feature was used.

Quantitative data were complemented with qualitative data collected through the poststudy interview. The aim of the interview was to produce contextual real-world knowledge about the behaviors and social structures of the participants, and their experiences with the MAAN app in their daily lives. The interview was conducted with parents or specialists of the participants and was regarding social and communication skills development.

#### *Quantitative Analysis*

The first questionnaire used was the SELSA-S questionnaire. The full version of the SELSA questionnaire is a 37-item self-report measure of romantic, social, and family loneliness. It was administered to young adults with ASD prestudy and poststudy [29,30]. Moreover, the Program for Education and Enrichment of Relational Skills mainly used it in most of their studies [21] to track the social and loneliness aspects of their participants. A study showed that the SELSA-S questionnaire, which consists of 15 items, is a psychometrically reliable and valid alternative to the full version of SELSA, and requires less time for participants [32]. There are 5 items referring to the family loneliness subscale, 6 items referring to the romantic subscale, and 4 items referring to social loneliness [33]. All items are answered on a 7-point Likert scale. This study used only the social and family subscales, since emotional change was not in the interest of this study. The SELSA-S questionnaire was used prestudy and poststudy, and the results were compared to find the rate of change.

The Wilcoxon signed-rank test is a nonparametric test that is particularly suitable for examining the difference between pretest and posttest measures in a small sample [34], and it was used in this study. Statistical analysis was performed using IBM SPSS software (IBM Corp), and a $P$ value <.05 was considered statistically significant.

The second questionnaire was the uMARS questionnaire that was coded following the scoring guides provided by the questionnaire developers [22]. The questionnaire consists of the following 3 sections: app quality, app subjective quality, and perceived impact. Moreover, the rating is scored out of 5, based on the scoring criteria given by the questionnaire developers. This questionnaire was used poststudy to evaluate the app's quality. Descriptive statistical analysis of the uMARS rating was used to gain insights into the usability experience of the participants with ASD. Many studies have used the uMARS questionnaire to evaluate the usability of developed apps [35-37].

#### *Qualitative Analysis*

To analyze the interviews, open and axial coding phases from the grounded theory were used. Grounded theory is a systematic methodology extensively used in qualitative research. This approach aims to generate a substantive theory that links the investigated data to reality [38]. After conducting the poststudy interview with the participants, the interview was transcribed by the first author (MHFH). Open coding was then used to generate initial concepts from the data. This was followed by axial coding to establish connections between different concepts and categories [39].

Prestudy and poststudy interviews were conducted. The prestudy interview was conducted with 5 experts, 2 of whom worked in special educational programs (specialist and educational and ABA/applied verbal behavior consultant, and Information and Communications Technology access expert in educational programs). The other 3 worked in directing centers for adults with ASD (director of the Rehabilitation and Inclusion Office, manager of the Severe Difficulties Department, and president of the Disability Association). The interviews were conducted for the purpose of obtaining insights into the current situation and existing technologies (state-of-the-art), and obtaining feedback on the study design. The poststudy interview was conducted with all participants' parents or caregivers to gather information about the context of use. The interview included questions about the overall app experience in terms of usability, functionality, and challenges encountered while using the app, as well as suggestions for further enhancement.

## Results

### Demographic Information

Among the 7 adults with ASD, 5 contacted their parents through the app, 1 contacted both a parent and specialist (ABA specialist and teacher with 25 years of experience), and 1 contacted only a specialist (speech and language therapist with 5 years of experience) (Table 1). Moreover, there was diversity in



XSL•FO
RenderX

educational background among the adults with ASD. This diversity in educational background was related to the severity of their conditions. Most of the parents and caregivers interacted with the participants for less than 6 hours per day. Based on the information gathered from the participants' parents, 2 participants were diagnosed with high-functioning ASD (participants 3 and 4), 3 were diagnosed with medium-functioning ASD (participants 5, 6, and 7), and 2 were diagnosed with low-functioning ASD and up to severe ASD (participants 1 and 2). Participant information is presented in Table 2.

**Table 1.** Demographic information.

| Characteristic | Adults with autism spectrum disorder (N=7), n | Parents (N=6), n | Specialists (N=2), n |
|---|---|---|---|
| **Gender** | | | |
| Female | 0 | 5 | 2 |
| Male | 7 | 1 | 0 |
| **Age (years)** | | | |
| 18-21 | 4 | 0 | 0 |
| 21-26 | 1 | 1 | 0 |
| >26 | 2 | 5 | 2 |

**Table 2.** Characteristics of the participants with autism spectrum disorder.

| Characteristic described by a parent or specialist | Participant |
|---|---|
| Completely nonverbal: Cannot speak but understands spoken conversation | Participants 1, 2, 3, and 5 |
| Verbal: Can greet | Participant 4 |
| Nonverbal: Cannot speak a lot and only greets, but understands spoken conversation | Participants 6 and 7 |

## Participant Engagement

The user study evaluation lasted 16 weeks. However, not all participants completed this duration due to this project's time limitation. Three participants completed the 16-week duration, whereas 2 participants used the app for 13 weeks, 1 participant used it for 11 weeks, and 1 participant used it for 10 weeks. The full duration of the study was tailored to each participant according to their enrolled duration. In this study, the duration was defined as the period from the start day after enrollment and training to the day when the poststudy interview was conducted. Participants' daily activities were monitored through interaction log data. Table 3 represents a summary of the average number of days the app was accessed by each participant per week. Although it was difficult to determine who initiated a conversation, since it could last or hold for a while, initiation was considered when the participant began the chat each day. Based on responses in the poststudy interview, 2 participants initiated some of the conversations, especially when waking up in the morning, needing something to eat, or needing to go out (requesting a need); 2 participants partially initiated conversations with some encouragement from their parents or caregivers; and 3 participants did not initiate any conversations.

**Table 3.** Average number of days per week the app was accessed by each participant.

| Participant | Value (days/week)[a], mean (SD) |
|---|---|
| 1 | 5.00 (1.826) |
| 2 | 2.37 (1.821) |
| 3 | 2.87 (1.893) |
| 4 | 4.62 (1.981) |
| 5 | 1.69 (1.109) |
| 6 | 3.36 (2.203) |
| 7 | 2.10 (1.287) |

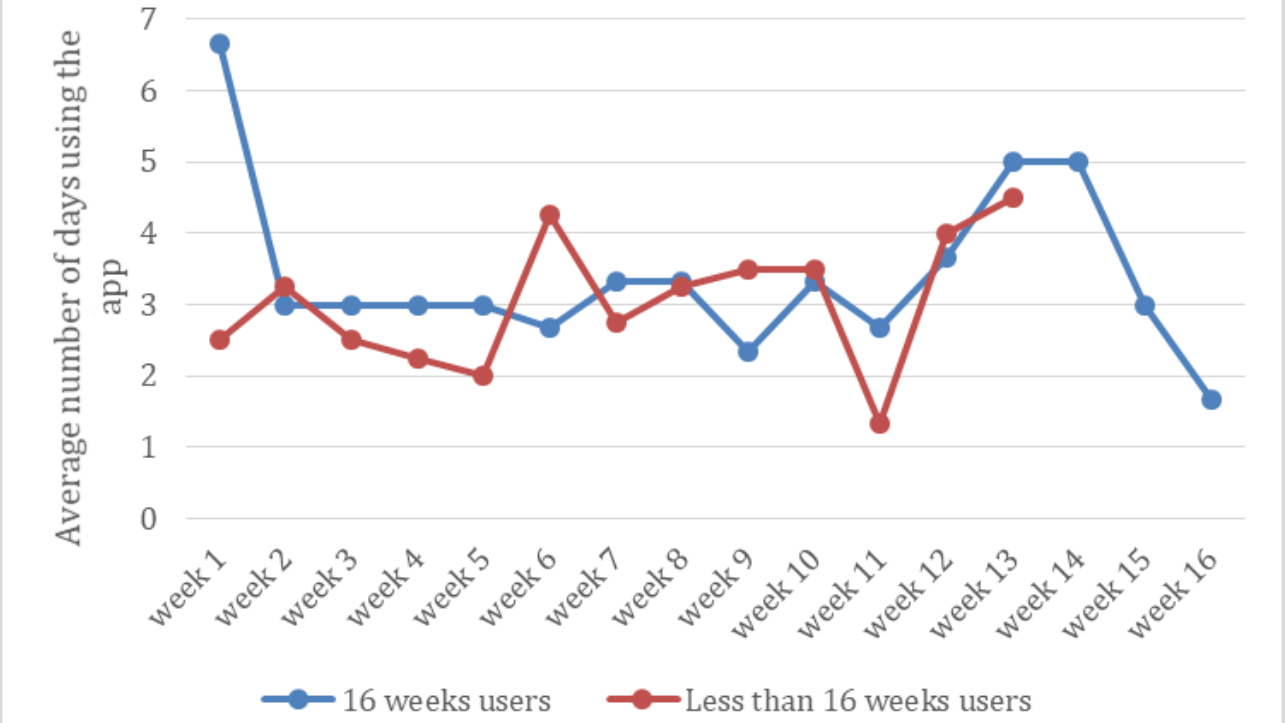
[a]The overall maximum and minimum values were 5 and 1.69 days/week, respectively, and the overall mean value was 3.14 (SD 1.25) days/week.

Based on the collected data, diversity in participant engagement was noticed. The participants were categorized into the following 2 groups: participants who used the app for 16 weeks and participants who used the app for less than 16 weeks. Figure 3 shows the average participant engagement for the 2 groups.



XSL•FO
RenderX

**Figure 3.** Participant engagement in the app.



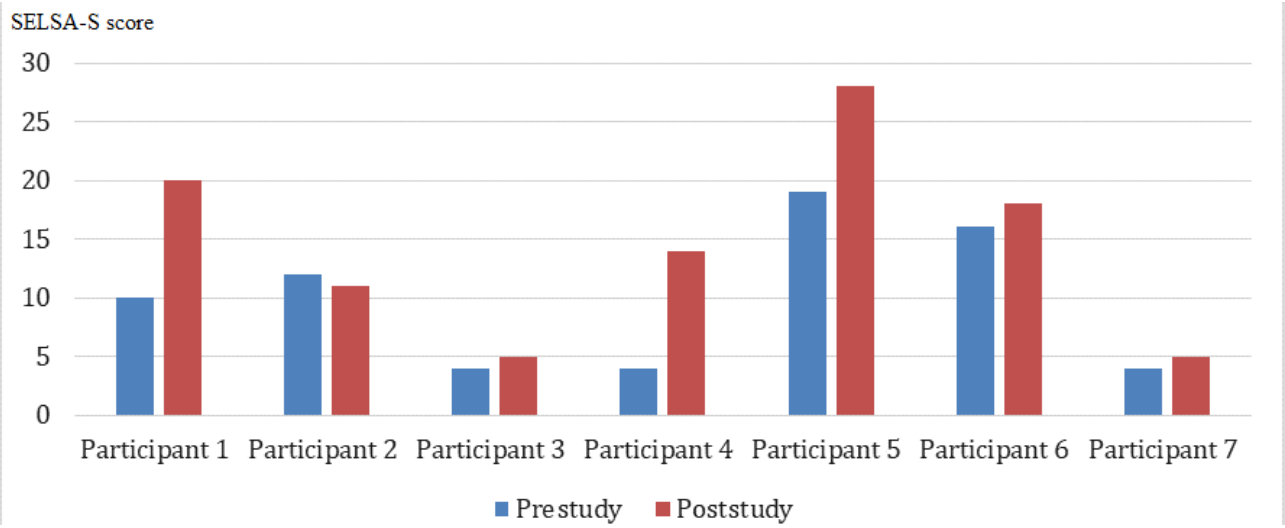

## Testing and Analysis Tools

### *Quantitative Analysis*

In this study, the social and family loneliness subscales were considered. It was evident from the collected data that the family loneliness subscale did not change from prestudy to poststudy. However, the social loneliness subscale showed a difference in all participants (Figure 4).

The use of each app feature differed from one participant to another. However, it was evident that when the app features were frequently used, the SELSA-S rate of change tended to be higher (Figure 5). For example, participant 1 used TTS more often than all other participants, and thus, the SELSA-S rate of change of this participant was the highest. Participant 7 used the app features less frequently and did not get engaged with the app when compared with the other participants. The SELSA-S rate of change for this participant was low.

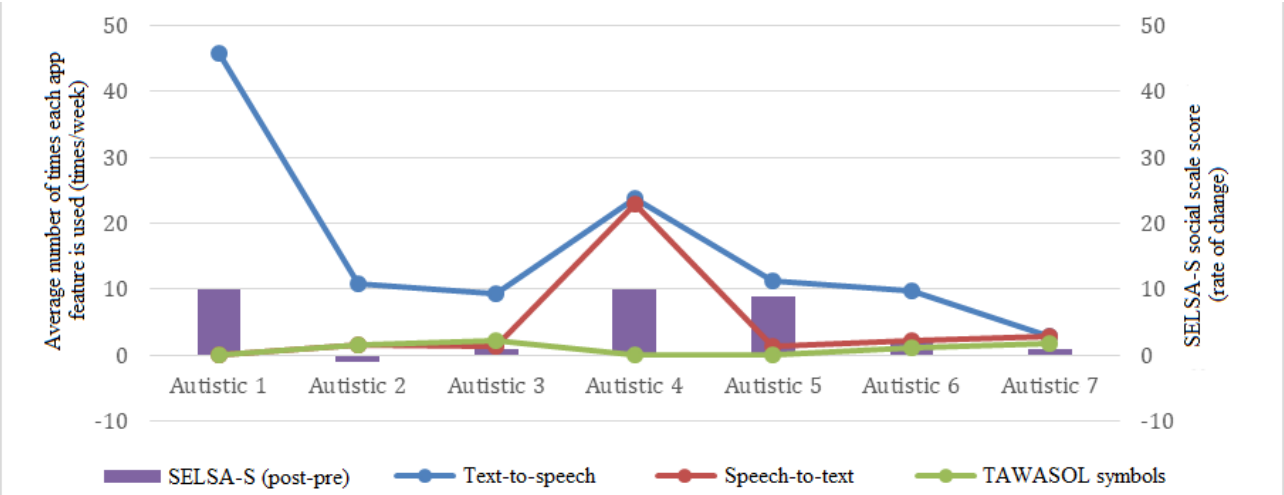
**Figure 4.** Prestudy and poststudy SELSA-S questionnaire social scale results. SELSA-S: short version of the Social and Emotional Loneliness Scale for Adults.






XSL•FO
RenderX

**Figure 5.** Average number of times each app feature is used per week with respect to the rate of change of the SELSA-S questionnaire result. SELSA-S: short version of the Social and Emotional Loneliness Scale for Adults.




Regarding uMARS results, first, the app quality mean score was calculated by finding the mean of the ratings for the subsections engagement, functionality, esthetics, and information, and then calculating the average of the 4 means. The mean value was 3.62 (SD 0.65). Second, app subjective quality was rated based on the mean score of the related questions, which was 3.72 (SD 1.12). Third, the questionnaire's perceived impact section consists of 6 questions that rate the impact of the app on the participants' knowledge, attitude, and intention regarding the targeted health behavior. Moreover, the rating was based on the mean score of the related questions, which was 3.54 (SD 0.98).

Point-biserial correlation was employed to answer the first hypothesis. The number of times each app feature was used was considered a dichotomous variable, since it is either clicked or not [40,41]. Therefore, point-biserial correlation, which is a special case of Pearson correlation, was used to measure the strength of the association between the SELSA-S rate of change (post-pre) and each app feature. The strength of the correlation was assessed as follows: week correlation if $0.1<|r|<0.3$, moderate correlation if $0.3<|r|<0.5$ and strong correlation if $0.5<|r|$. The results showed that the SELSA-S rate of change was strongly related to the TTS feature ($r$=0.708, $P$=.38) and to the TAWASOL symbols feature ($r$=−0.917, $P$=.002). On the other hand, the SELSA-S rate of change was moderately related to the STT feature ($r$=0.428, $P$=.17). Therefore, usage of TTS, STT, and CS (TAWASOL symbols) can enhance the social and communication skills of adults with ASD. For the second hypothesis, the Wilcoxon signed-rank test was used to determine whether the difference between the prestudy and poststudy results in the social loneliness scale was significant. It indicated that poststudy SELSA-S results were statistically significantly higher than prestudy SELSA-S results ($z$=−2.047; $P$=.04). Thus, the null hypothesis was rejected. Therefore, usage of the app can reduce social loneliness among adults with ASD.

### *Qualitative Analysis*

All the interviewees in the prestudy interviews agreed on the importance of the MAAN app as a unique tool that could be of great use. They stated that this innovative app would assist centers, parents, and caregivers in staying connected with adults with ASD. They also believed that MAAN has the potential to enhance the social and communication skills of adults with ASD. Moreover, they asserted that such an app could support the inclusion of adults with ASD into social, educational, and work settings, ultimately encouraging adults with ASD to use more social-based apps.

Overall, 6 parents and 1 specialist were interviewed in the poststudy interview. The interview focused on their experiences with the app, the challenges faced, and the recommended modifications. The open and axial coding phases of the grounded theory [39] were used to analyze the transcribed interview. The analysis highlighted the importance of the app in enabling communication between adults with ASD and other individuals. Most interviewees discussed the relevance of the app and its novelty. Five categories emerged from the analysis, which were app support, feature relevance, user interface design, overall feedback, and recommendations. Each category had subcategories and samples of excerpts from the transcribed data describing the experiences of the participants with ASD while using the app (Table 4).



XSL•FO
RenderX

**Table 4.** Categories and subcategories that emerged from the poststudy interview.

| Category and subcategory | Description | Example quotes |
|---|---|---|
| **App support** | Technical and customized support to access the needed information without problems or errors | |
| Learnability | Interactive app features, which aid participants with ASD[a] to learn quickly | *...He was able to understand how to use it after I trained him...* [Parent #4] |
| Efficiency (accessibility) | Accessing the features of the app (easy, medium, or hard) | *...He accepted the application very smoothly and enjoyed using it...* [Parent #4] |
| Technical issues | Bugs or related technical issues | *...He tried to use recording, but it was a little hard for him to use it...* [Parent #5] |
| **Feature relevance** | Feedback of the participants on the app features | |
| Text-to-speech | Feedback on the text-to-speech feature | *…it was an amazing experience for him to listen to the text, while this feature is not available in any other platform even WhatsApp...* [Parent #1] |
| Speech-to-text | Feedback on the speech-to-text feature | *…Speech to text is a very good feature I would say especially for autistic case…* [Parent #6] |
| Arabic issues | Participants talk about Arabic issues related to the app | *…The Arabic language was a little annoying since text-to-speech and speech-to-text transcriptions were robotic and need more advancements…* [Parent #2] |
| TAWASOL symbols | Feedback on TAWASOL symbols | *…For him, both TAWASOL symbols and voice messaging are the main features that helped him…* [Parent #2] |
| **User interface design** | Visual appearance of the app, such as the arrangement of content, color schemes, icons, and font sizes | |
| Color | Feedback on the color changing tool of the app | *…He didn't change the application color nor the chat bubble since he doesn't have any problem with colors, but other autistic individuals can use this feature to help them using the application more by changing the color with what is suitable for them…* [Parent #2] |
| Design consistency | The participant likes texting and using the app like other messaging apps | *…He knows [refereeing to the participants with ASD] how to type and use it in other application, then he liked the chatting (Texting) part more…* [Parent #1] |
| **Overall feedback** | Feedback about the app in general regarding the innovative idea and the need of it | |
| Positive feedback | Positive feedback on the app | *…Really good application for chatting like I'm very happy about the way it sends messages in different format text-to-speech, speech-to-text, and Tawasol symbols, so I was introducing it to him as a chatting app itself and he liked it* [Caregiver #1] |
| Novelty | Feedback on the novelty of the app compared to existing messaging apps | *…I'm just saying because it's a very good with its unique features what I really found very interesting since I can't find these features in other similar applications such as WhatsApp…* [Parent #1] |
| **Recommendations** | Suggestions of parents and specialists on how to improve the app | |
| Serious games | Suggestions to enhance the app with game elements to teach specific skills and knowledge | *…I would like to suggest including games where autistic adults could learn the TAWASOL symbols and be able to easily construct sentences…* [Parent #3] |
| Privacy protection | Discussing the next version based on a friend request feature in order to protect users from strangers | *…Caregivers should be able to accept friends, and this is because we need to protect them from unknown people…* [Parent #3] |

[a]ASD: autism spectrum disorder.



XSL•FO
RenderX

## Discussion

### Summary

In this study, MAAN, a mobile messaging app, was developed and evaluated in 7 adults with ASD over 10 to 16 weeks. A pre-post study was also conducted with experts and parents on the interface and functionality design of the app. The app has additional features (TTS, STT, and CS) when compared with existing messaging apps. MAAN is designed to support adults with ASD when communicating with other individuals via text messaging in both Arabic and English. It offers the ability for adults with ASD to read or listen to text messages and then reply. The results supported the hypotheses and are encouraging for further work in the future.

### Comparison With Related Work

Very few messaging apps that support social and communication skills in individuals with ASD are discussed in the literature. TalkingBoogies [42], an app introduced in 2020 by a research team in Korea, aimed to actively assist caregivers when developing AAC-led communication with children having ASD. It comprises the following 2 iOS apps: TalkingBoogies-AAC for child caregiver communication and TalkingBoogies-Coach for caregiver collaboration. TalkingBoogies-AAC has several features, such as the use of Ewha [43], an AAC symbol system, and TTS. The team evaluated the developed app in 4 children with ASD and 3 teachers with at least 6 months of experience with AAC from a local special education school. The study concluded that such an app can prompt the learnability of children when constructing sentences. Prior to this study, De Leo et al [44] built and evaluated a Windows-based mobile app named "PixTalk." The PixTalk system employs the PECS intervention and is made up of the following 2 modules: PixTalk smartphones, which enables children with ASD to search and select images in order to express their needs and feelings, and PixTalk website, which allows caregivers and teachers to add different images to a child's smartphone. By conducting a case study involving 3 children with ASD and their teachers, this research asserted the importance of computer-assisted instructions as an intervention to motivate and engage learners with ASD. Proloquo2Go [45] and Proloquo4Text [46] are popular communication apps that use symbols and the TTS feature. Both apps are commercially available and widely used in school settings where English is the main instructional language. Nowadays, virtual communication platforms, such as social media networks and messaging apps, are becoming more integrated in most people's lives [47]. However, from an accessibility viewpoint, the use of these platforms is neither practical nor accessible for the majority of individuals with ASD and especially for the nonverbal population. Some of the existing work involves AACrobat [48]. The authors designed this mobile app to help neuromuscular disease users communicate with others using their eyes. Although AACrobat uses the TTS feature, this apps is not helpful for individuals with ASD since most of them cannot fix their eyes in one direction to type from a keyboard or do not have the ability to generate a sentence. Currently, the most popular communication platform is WhatsApp [47], which is mainly used as a messaging app. Despite all the important updates and features that WhatsApp provides, it is still not accessible for individuals with ASD, as most of the nonverbal population struggles to read, write, or communicate using basic texting features. To the best of our knowledge, previous studies have attempted to support adults with ASD in constructing sentences for education and learning purposes, but no study has evaluated TTS, STT, and CS features, as well as social and family loneliness in adults with ASD. This paper presents the development and evaluation of the new messaging app MAAN, which means "together" in Arabic. MAAN employs TTS, STT, and CS to enhance the social and communication skills of adults with ASD. This is the first study to consider these features in a mobile messaging app designed to support communication among adults with ASD, both in English and Arabic. The results showed that MAAN can enhance social and communication skills, especially through distance messaging, and can ease communication with peers through its special features (TTS, STT, and CS). Hence, this study showed the effectiveness of these features in a messaging app.

### Principal Results

The 5 experts interviewed in the prestudy phase emphasized the importance of MAAN as a unique tool in supporting adults with ASD. During the interview, they discussed the app's novelty and suggested several modifications, such as categorizing TAWASOL symbols and fixing the app's design. Moreover, they emphasized the support of the app for the inclusion of adults with ASD into social, educational, and work settings, ultimately encouraging adults with ASD to use more social-based apps. On the other hand, the poststudy interview with parents suggested that the app provides 3 features directed to verbal and nonverbal adults with ASD, and the TTS feature was the most preferable feature for both ASD and non-ASD participants. Using MAAN helped the participants with ASD interact more with their parents, especially when they were not near them. Besides, the participants with ASD were attached to the app, where they voiced their needs by sending a message to their parents. Support for the Arabic language in all 3 features made this app very appealing to the participants, considering that there is a lack of ACC apps that support the Arabic language. Moreover, the interviewees highlighted the importance of MAAN in enhancing the social and communication skills of different users, such as individuals with dyslexia. From the participants' engagements with the app and the SELSA-S questionnaire results, it can be deduced that more use of the app features was associated with higher SELSA-S questionnaire scores. Moreover, this was confirmed in the poststudy interviews where the interviewees highlighted the importance of the app features and their ability to increase the attention of adults with ASD, which can positively improve social loneliness. This development is noticeable with the positive engagement and enthusiasm that the participants with ASD exhibited when messaging their parents using the app features or when constructing a sentence by using TAWASOL symbols

### Limitations

Despite this study's contributions, some limitations, including the sample size, operating system platform, and choice of features, were noted. Due to the COVID-19 pandemic, most



XSL•FO
RenderX

adults with ASD were at home with their families and thus could not be easily reached. This led to a rather small sample. Another factor that contributed to the small sample was the absence of focused institutions for adults with ASD, since most of the current institutions serve children with ASD. A larger sample size could give more insights into the effects of the app and could increase the statistical power. Further, the message app was only available on the iOS platform. This availability limitation was due to the timeframe for developing the app. Lastly, other messaging app features, such as image messaging, video messaging, and voice and video calls, were excluded. The reason is to prompt adults with ASD to use AAC-based features rather than video or voice calls, which, in turn, can support their social and communication skills.

## Conclusion

The novelty of MAAN as a communication and social intervention app is its potential to support communication skills and social loneliness in adults with ASD. More importantly, it was possible to achieve this due to the inclusion of experts in the design and development of the app. Additionally, the poststudy evaluation by parents and specialists identified the uniqueness of the app, and how it could be enhanced and extended to other populations who also exhibit social and communication deficits, such as people with dyslexia. Future studies can consider a larger number of participants with ASD to replicate the findings and can extend this study to other clinical populations with social and communication deficits.

## Acknowledgments

We would like to express our gratitude to Hamad Bin Khalifah University (HBKU). We also gratefully acknowledge all the participants who took part in the study. Their dedication and commitment have had an enormous impact on the quality of this research. We would like to thank the administration of Step by Step Center (for Special Educational Needs) and Best Buddies Center in Qatar, and the Autism Learning Institute for Applied Behaviour Analysis (ALI for ABA Center; Dr Chafica Mansour Gharbieh) in Lebanon for allowing their teaching staff, specialists, and users to take part in the study.

## Conflicts of Interest

None declared.

## References

1. Salhia HO, Al-Nasser LA, Taher LS, Al-Khathaami AM, El-Metwally AA. Systemic review of the epidemiology of autism in Arab Gulf countries. Neurosciences (Riyadh) 2014 Oct;19(4):291-296 [FREE Full text] [Medline: 25274588]
2. Schiltz HK, McVey AJ, Dolan Wozniak B, Haendel AD, Stanley R, Arias A, et al. The role of loneliness as a mediator between autism features and mental health among autistic young adults. Autism 2021 Feb;25(2):545-555. [doi: 10.1177/1362361320967789] [Medline: 33126822]
3. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
4. Kogan MD, Vladutiu CJ, Schieve LA, Ghandour RM, Blumberg SJ, Zablotsky B, et al. The Prevalence of Parent-Reported Autism Spectrum Disorder Among US Children. Pediatrics 2018 Dec;142(6) [FREE Full text] [doi: 10.1542/peds.2017-4161] [Medline: 30478241]
5. Alshaban F, Aldosari M, Al-Shammari H, El-Hag S, Ghazal I, Tolefat M, et al. Prevalence and correlates of autism spectrum disorder in Qatar: a national study. J Child Psychol Psychiatry 2019 Dec;60(12):1254-1268 [FREE Full text] [doi: 10.1111/jcpp.13066] [Medline: 31069792]
6. Richa S, Khoury R, JRouhayem J, Chammay R, Kazour F, Bou Khalil R, et al. Estimating the prevalence of autism spectrum disorder in Lebanon. Encephale 2020 Dec;46(6):414-419. [doi: 10.1016/j.encep.2020.05.015] [Medline: 32928536]
7. Keenan M, Dillenburger K, Röttgers HR, Dounavi K, Jónsdóttir SL, Moderato P, et al. Autism and ABA: The Gulf Between North America and Europe. Rev J Autism Dev Disord 2014 Dec 30;2(2):167-183. [doi: 10.1007/s40489-014-0045-2]
8. Thiemann-Bourque K. Peer-Mediated AAC Instruction for Young Children with Autism and other Developmental Disabilities. Perspect Augment altern commun 2012 Dec;21(4):159-166 [FREE Full text] [doi: 10.1044/aac21.4.159] [Medline: 24392179]
9. Carr D, Felce J. "Brief report: increase in production of spoken words in some children with autism after PECS teaching to Phase III". J Autism Dev Disord 2007 Apr 19;37(4):780-787. [doi: 10.1007/s10803-006-0204-0] [Medline: 17048094]
10. Satcher D. Mental health: A report of the Surgeon General--Executive summary. Professional Psychology: Research and Practice 2000;31(1):5-13. [doi: 10.1037/0735-7028.31.1.5]
11. Zisk AH, Dalton E. Augmentative and Alternative Communication for Speaking Autistic Adults: Overview and Recommendations. Autism in Adulthood 2019 Jun;1(2):93-100. [doi: 10.1089/aut.2018.0007]
12. Wendt O. Research on the use of manual signs and graphic symbols in autism spectrum disorders: A systematic review. In: Mirenda P, Iacono T, editors. Autism Spectrum Disorders and AAC. Baltimore, MD: Paul H. Brookes; 2009:83-140.
13. McNaughton D, Light J. What We Write about When We Write About AAC: The Past 30 Years of Research and Future Directions. Augment Altern Commun 2015;31(4):261-270. [doi: 10.3109/07434618.2015.1099736] [Medline: 26490318]



XSL•FO
RenderX

14. Howlin P, Gordon RK, Pasco G, Wade A, Charman T. The effectiveness of Picture Exchange Communication System (PECS) training for teachers of children with autism: a pragmatic, group randomised controlled trial. J Child Psychol Psychiatry 2007 May;48(5):473-481. [doi: 10.1111/j.1469-7610.2006.01707.x] [Medline: 17501728]
15. Nunes DRP. Augmentative and Alternative Communication for Autism Spectrum Disorder: A Review of an Emerging Practice. International Journal of Behavior Analysis & Autism Spectrum Disorders 2015;1:97-112 [FREE Full text]
16. Carr D, Felce J. The effects of PECS teaching to Phase III on the communicative interactions between children with autism and their teachers. J Autism Dev Disord 2007 Apr;37(4):724-737. [doi: 10.1007/s10803-006-0203-1] [Medline: 17006780]
17. Valencia K, Rusu C, Quiñones D, Jamet E. The Impact of Technology on People with Autism Spectrum Disorder: A Systematic Literature Review. Sensors (Basel) 2019 Oct 16;19(20):4485 [FREE Full text] [doi: 10.3390/s19204485] [Medline: 31623200]
18. Yee HSS. Mobile technology for children with Autism Spectrum Disorder: Major trends and issues. 2012 Presented at: 2012 IEEE Symposium on E-Learning, E-Management and E-Services; October 21-24, 2012; Kuala Lumpur, Malaysia p. 1-5. [doi: 10.1109/IS3e.2012.6414954]
19. Wojciechowski A, Al-Musawi R. Assisstive technology application for enhancing social and language skills of young children with autism. Multimed Tools Appl 2016 Oct 7;76(4):5419-5439. [doi: 10.1007/s11042-016-3995-9]
20. Boyd TK, Hart Barnett JE, More CM. Evaluating iPad Technology for Enhancing Communication Skills of Children With Autism Spectrum Disorders. Intervention in School and Clinic 2015 Apr 06;51(1):19-27. [doi: 10.1177/1053451215577476]
21. Laugeson EA, Gantman A, Kapp SK, Orenski K, Ellingsen R. A Randomized Controlled Trial to Improve Social Skills in Young Adults with Autism Spectrum Disorder: The UCLA PEERS(®) Program. J Autism Dev Disord 2015 Dec;45(12):3978-3989. [doi: 10.1007/s10803-015-2504-8] [Medline: 26109247]
22. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
23. Cunningham T, McNaughtan H. Text-to-Speech (TTS). Reading Rockets. URL: https://www.readingrockets.org/article/text-speech-tts [accessed 2021-02-21]
24. Shadiev R, Hwang WY, Chen NS, Huang YM. Review of Speech-to-Text Recognition Technology for Enhancing Learning. Educational Technology & Society 2014;17(4):65-84 [FREE Full text]
25. Elsheikh A, Zeinon N. 2019 Presented at: 7th International conference on ICT & Accessibility (ICTA); December 13-15, 2019; Hammamet, Tunisia p. 1-5. [doi: 10.1109/ICTA49490.2019.9144883]
26. Sofian N, Sobri HA, Sarlan A. Multimedia Elements in Designing Mobile App Interface for Autistic Children: Proxy User Perspective. 2021 Presented at: 2021 International Conference on Computer & Information Sciences (ICCOINS); July 13-15, 2021; Kuching, Malaysia p. 96-101. [doi: 10.1109/ICCOINS49721.2021.9497180]
27. Ntalindwa T, Nduwingoma M, Karangwa E, Rashid Soron T, Uworwabayeho A, Uwineza A. Development of a Mobile App to Improve Numeracy Skills of Children With Autism Spectrum Disorder: Participatory Design and Usability Study. JMIR Pediatr Parent 2021 Aug 31;4(3):e21471 [FREE Full text] [doi: 10.2196/21471] [Medline: 34463629]
28. Naderifar M, Goli H, Ghaljaie F. Snowball Sampling: A Purposeful Method of Sampling in Qualitative Research. Strides Dev Med Educ 2017 Sep 30;14(3). [doi: 10.5812/sdme.67670]
29. Gantman A, Kapp SK, Orenski K, Laugeson EA. Social skills training for young adults with high-functioning autism spectrum disorders: a randomized controlled pilot study. J Autism Dev Disord 2012 Jun 14;42(6):1094-1103. [doi: 10.1007/s10803-011-1350-6] [Medline: 21915740]
30. McVey AJ, Dolan BK, Willar KS, Pleiss S, Karst JS, Casnar CL, et al. A Replication and Extension of the PEERS® for Young Adults Social Skills Intervention: Examining Effects on Social Skills and Social Anxiety in Young Adults with Autism Spectrum Disorder. J Autism Dev Disord 2016 Dec;46(12):3739-3754 [FREE Full text] [doi: 10.1007/s10803-016-2911-5] [Medline: 27628940]
31. LeBeau K, Huey LG, Hart M. Assessing the Quality of Mobile Apps Used by Occupational Therapists: Evaluation Using the User Version of the Mobile Application Rating Scale. JMIR Mhealth Uhealth 2019 May 01;7(5):e13019 [FREE Full text] [doi: 10.2196/13019] [Medline: 31066712]
32. Cramer KM, Ofosu HB, Barry JE. An abbreviated form of the social and emotional loneliness scale for adults (SELSA). Personality and Individual Differences 2000 Jun;28(6):1125-1131. [doi: 10.1016/s0191-8869(99)00162-2]
33. Pollet TV, Saxton TK, Mitchell M. Measurement equivalence between men and women in the abbreviated Social and Emotional Loneliness Scale for Adults (SELSA). Interpersona 2018 Dec 21;12(2):283-292. [doi: 10.5964/ijpr.v12i2.316]
34. Mier N, Tanguma J, Millard AV, Villarreal EK, Alen M, Ory MG. A pilot walking program for Mexican-American women living in colonias at the border. Am J Health Promot 2011;25(3):172-175. [doi: 10.4278/ajhp.090325-ARB-115] [Medline: 21192746]
35. Peek J, Hay K, Hughes P, Kostellar A, Kumar S, Bhikoo Z, et al. Feasibility and Acceptability of a Smoking Cessation Smartphone App (My QuitBuddy) in Older Persons: Pilot Randomized Controlled Trial. JMIR Form Res 2021 Apr 14;5(4):e24976 [FREE Full text] [doi: 10.2196/24976] [Medline: 33851923]




XSL•FO
RenderX

36. Herbuela VRDM, Karita T, Carvajal TM, Ho HT, Lorena JMO, Regalado RA, et al. Early Detection of Dengue Fever Outbreaks Using a Surveillance App (Mozzify): Cross-sectional Mixed Methods Usability Study. JMIR Public Health Surveill 2021 Mar 1;7(3):e19034 [FREE Full text] [doi: 10.2196/19034] [Medline: 33646128]
37. Chen R, Santo K, Wong G, Sohn W, Spallek H, Chow C, et al. Mobile Apps for Dental Caries Prevention: Systematic Search and Quality Evaluation. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e19958 [FREE Full text] [doi: 10.2196/19958] [Medline: 33439141]
38. Urquhart C. Grounded Theory for Qualitative Research: A Practical Guide. New York, NY: SAGE Publications Ltd; 2013.
39. Corbin J, Strauss A. Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory. New York, NY: SAGE Publications, Inc; 2008.
40. MacKenzie IS. Human-Computer Interaction: An Empirical Research Perspective. Burlington, MA: Morgan Kaufmann; 2013.
41. Farzan R, DiMicco J, Millen D, Dugan C, Geyer W, Brownholtz E. Results from deploying a participation incentive mechanism within the enterprise. In: CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: Conference on Human Factors in Computing Systems; April 5-10, 2008; Florence, Italy p. 563-572. [doi: 10.1145/1357054.1357145]
42. Shin D, Song J, Song S, Park J, Lee J, Jun S. TalkingBoogie: Collaborative Mobile AAC System for Non-verbal Children with Developmental Disabilities and Their Caregivers. In: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI p. 1-13. [doi: 10.1145/3313831.3376154]
43. Park E, Kim Y, Hong K, Yeon S, Kim K, Lim J. Development of Korean Ewha-AAC Symbols: Validity of Vocabulary and Graphic Symbols. aac 2016 Dec 31;4(2):19. [doi: 10.14818/aac.2016.12.4.2.19]
44. De Leo G, Gonzales CH, Battagiri P, Leroy G. A smart-phone application and a companion website for the improvement of the communication skills of children with autism: clinical rationale, technical development and preliminary results. J Med Syst 2011 Aug 2;35(4):703-711. [doi: 10.1007/s10916-009-9407-1] [Medline: 20703781]
45. Sennott S, Bowker A. Autism, AAC, and Proloquo2Go. Perspect Augment Altern Commun 2009 Dec;18(4):137-145. [doi: 10.1044/aac18.4.137]
46. Lund SK, Quach W, Weissling K, McKelvey M, Dietz A. Assessment With Children Who Need Augmentative and Alternative Communication (AAC): Clinical Decisions of AAC Specialists. LSHSS 2017 Jan;48(1):56-68. [doi: 10.1044/2016_lshss-15-0086]
47. SPECTRM. 2020. URL: https://tinyurl.com/yckuycf3 [accessed 2021-02-24]
48. Fiannaca A, Paradiso A, Shah M, Morris MR. AACrobat: Using Mobile Devices to Lower Communication Barriers and Provide Autonomy with Gaze-Based AAC. In: CSCW '17: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing; February 25-March 1, 2017; Portland, OR p. 683-695. [doi: 10.1145/2998181.2998215]

## Abbreviations

**AAC:** augmentative and alternative communication
**ABA:** applied behavioral analysis
**ASD:** autism spectrum disorder
**CS:** communication symbols
**PECS:** picture exchange communication system
**SELSA-S:** short version of the Social and Emotional Loneliness Scale for Adults
**STT:** speech-to-text
**TTS:** text-to-speech
**uMARS:** Mobile Application Rating Scale: user version

*Edited by G Eysenbach; submitted 24.08.21; peer-reviewed by 신 , MDG Pimentel; comments to author 03.10.21; revised version received 08.11.21; accepted 17.11.21; published 07.12.21.*

*Please cite as:*
*Hijab MHF, Al-Thani D, Banire B*
*A Multimodal Messaging App (MAAN) for Adults With Autism Spectrum Disorder: Mixed Methods Evaluation Study*
*JMIR Form Res 2021;5(12):e33123*
*URL: https://formative.jmir.org/2021/12/e33123*
*doi:10.2196/33123*
*PMID:34878998*



XSL•FO
RenderX

©Mohamad Hassan Fadi Hijab, Dena Al-Thani, Bilikis Banire. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Assessing Health Data Security Risks in Global Health Partnerships: Development of a Conceptual Framework

Juan Espinoza[1,2], MD; Abu Taher Sikder[3], BA; James Dickhoner[2,3], MD; Thomas Lee[2,4], MD

[1]Department of Pediatrics, Children's Hospital Los Angeles, Los Angeles, CA, United States
[2]Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
[3]Innovation Studio, Children's Hospital Los Angeles, Los Angeles, CA, United States
[4]Department of Surgery, Children's Hospital Los Angeles, Los Angeles, CA, United States

**Corresponding Author:**
Juan Espinoza, MD
Department of Pediatrics
Children's Hospital Los Angeles
4650 Sunset Blvd
Los Angeles, CA, 90006
United States
Phone: 1 3233612721
Email: jespinoza@chla.usc.edu

## Abstract

**Background:** Health care databases contain a wealth of information that can be used to develop programs and mature health care systems. There is concern that the sensitive nature of health data (eg, ethnicity, reproductive health, sexually transmitted infections, and lifestyle information) can have significant impact on individuals if misused, particularly among vulnerable and marginalized populations. As academic institutions, nongovernmental organizations, and international agencies begin to collaborate with low- and middle-income countries to develop and deploy health information technology (HIT), it is important to understand the technical and practical security implications of these initiatives.

**Objective:** Our aim is to develop a conceptual framework for risk stratification of global health data partnerships and HIT projects. In addition to identifying key conceptual domains, we map each domain to a variety of publicly available indices that could be used to inform a quantitative model.

**Methods:** We conducted an overview of the literature to identify relevant publications, position statements, white papers, and reports. The research team reviewed all sources and used the framework method and conceptual framework analysis to name and categorize key concepts, integrate them into domains, and synthesize them into an overarching conceptual framework. Once key domains were identified, public international data sources were searched for relevant structured indices to generate quantitative counterparts.

**Results:** We identified 5 key domains to inform our conceptual framework: State of HIT, Economics of Health Care, Demographics and Equity, Societal Freedom and Safety, and Partnership and Trust. Each of these domains was mapped to a number of structured indices.

**Conclusions:** There is a complex relationship among the legal, economic, and social domains of health care, which affects the state of HIT in low- and middle-income countries and associated data security risks. The strength of partnership and trust among collaborating organizations is an important moderating factor. Additional work is needed to formalize the assessment of partnership and trust and to develop a quantitative model of the conceptual framework that can help support organizational decision-making.

***(JMIR Form Res 2021;5(12):e25833)*** doi:10.2196/25833

**KEYWORDS**

health information technology; low- and middle-income countries; low income; conceptual framework analysis; framework method; data security; decision-making; database; information use; misuse; global health; security



XSL•FO
RenderX

## Introduction

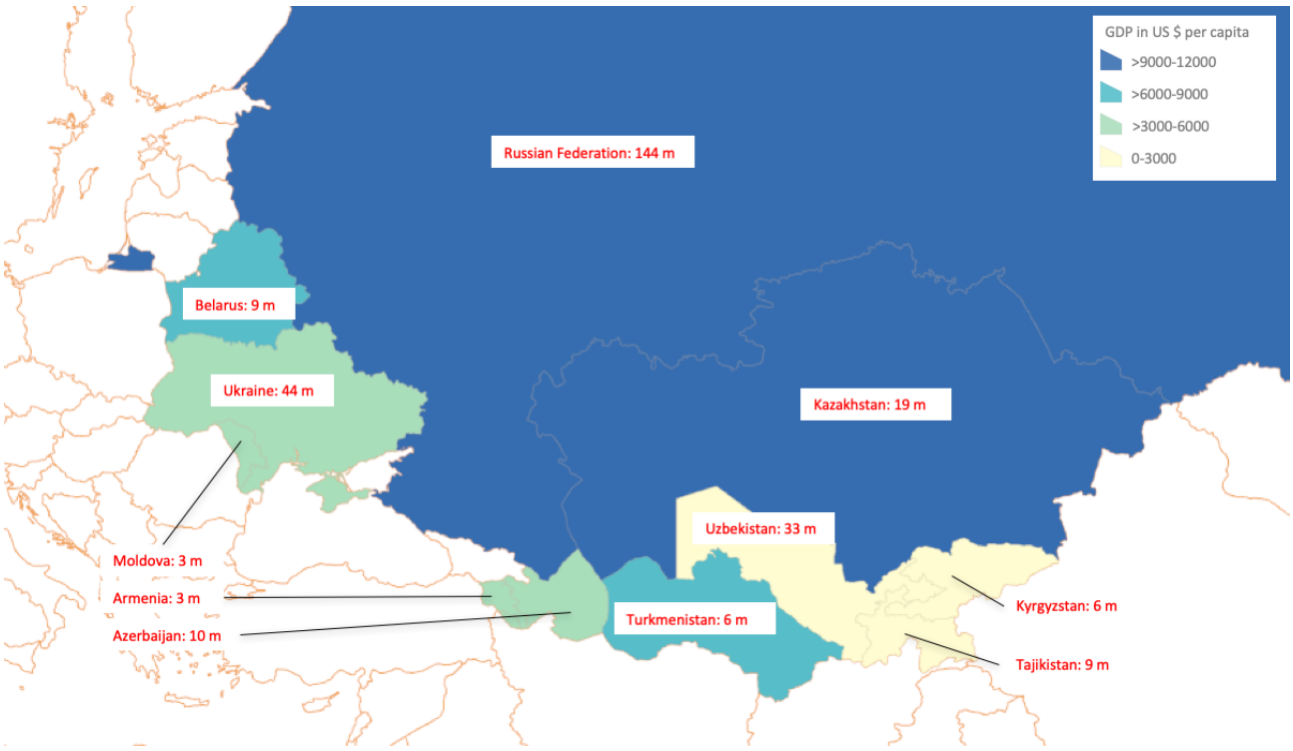
### Background

Health information technology (HIT) refers to electronic health records (EHRs), patient portals and other software platforms, public health databases, hardware devices, and technology systems, which contain a wealth of information used for patient care and resource allocation. According to a 2016 World Health Organization (WHO) global survey on eHealth, the adoption of EHR systems had increased by 46% in the previous 5 years [1]. Increasing numbers of low- and middle-income countries (LMICs) are implementing HIT as they continue developing their health care infrastructure [2,3]. In many cases, these systems are being implemented in collaboration with foreign academic institutions, health care systems, and nonprofit or research organizations [4,5]. However, there are unique organizational, technical, functional, educational, and ethical challenges that require meticulous consideration—especially with respect to their security implications [6].

In many LMICs, legal, regulatory, and technical frameworks around HIT are undeveloped [6]. Furthermore, the sensitive nature of health data (eg, ethnicity, reproductive health, sexually transmitted infections, and lifestyle information) can have significant impact on individuals if misused [7]. The combination of developing legal frameworks and decreased ability of public institutions to protect individuals may create a particularly vulnerable environment for HIT and health data. A framework to understand and stratify the risk associated with HIT may be beneficial to organizations engaging in global health partnerships that generate significant amounts of health data.

Our research team has been focused on international clinical and research partnerships in Armenia. As we have engaged in the process of designing and deploying a safe, scalable health data platform in Armenia, other countries in the region have expressed interest in implementing similar systems. This geopolitical region is home to the Commonwealth of Independent States (CIS), an intergovernmental organization of 11 post-Soviet countries: Armenia, Azerbaijan, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan (Figure 1) [8]. These countries inherited the Soviet Union's Semashko health system and many of its flaws in transitioning to a modern primary care model, including premature mortality, variable quality of care, poor noncommunicable disease management, and high out-of-pocket payments [9-11]. Rising social and economic challenges such as inequality and the cost of funding public health are also significant concerns [9]. To address these concerns, CIS member states have embarked on health care reform efforts to improve their health care systems, including deploying HIT [9,12].

**Figure 1.** A map of Commonwealth of Independent States countries located in Eastern Europe and Central Asia. The color gradient shows 2019 gross domestic product per capita in US dollars. The population of the country (in millions) is noted next to the name. GDP: gross domestic product.




### Need for a Framework

As our organization considers new partnerships among other CIS members, the need for a framework to understand potential HIT security risks has become more pressing. Given the lack of an existing risk stratification framework to assess and consider the security vulnerabilities with implementing HIT in LMICs, we decided to create our own. In this paper we present a conceptual approach to developing such a framework and propose a variety of indices that could be leveraged to inform each subdomain. Finally, we outline our proposed next steps to



XSL•FO
RenderX

seek consensus and finalize quantitative and qualitative versions of the framework.

## Methods

### Literature Review

We reviewed the literature to identify relevant publications by searching PubMed, Ovid MEDLINE, Google Scholar, and Google using the search terms presented in Textbox 1. All articles were identified as relevant by at least 2 authors (JE, ATS, or JD). References from relevant articles were also reviewed. Original research, reviews, editorials and commentaries, position statements, white papers, and industry and nongovernmental organization reports were included. Finally, the websites of international agencies such as the WHO, the United Nations, and the World Bank were reviewed for relevant data sources and publications.

**Textbox 1.** Literature search terms by concept type.

**Search terms**

- Geographic terms
  - Armenia
  - Europe
  - Asia
  - Soviet
  - Commonwealth of Independent States (CIS)
  - Low- and middle-income countries (LMIC)
  - Developing countries
  - Global health
  - International
- Technology terms
  - Data
  - Health data
  - Patient data
  - Information
  - Health information
  - Patient information
  - Electronic health records (EHR)
  - Electronic medical records (EMR)
  - Health information technology (HIT)
  - Personal health records
  - Database
- Theme terms
  - Security
  - Risk
  - Partnership
  - Breach
  - Trust
  - Safety
  - Health system
  - Health policy
  - Legal

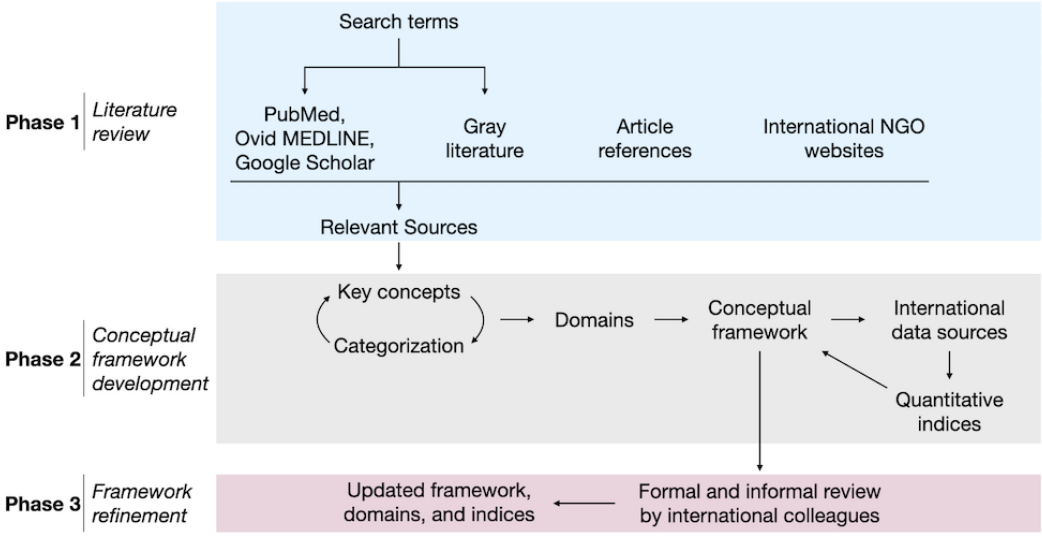


XSL•FO
RenderX

### Conceptual Framework Development

A conceptual framework is a "network of...interlinked concepts that allow for a comprehensive understanding of a phenomenon or phenomena [13]." Our conceptual framework development process borrowed iterative techniques from both the framework method [14] and conceptual framework analysis [13]. First, all literature sources were extensively reviewed for key concepts. By categorizing and recategorizing these key concepts iteratively as more literature was reviewed, the team conceptualized them into domains that were then synthesized to develop an overarching conceptual framework using the coding methodology outlined by Gale et al [14]. Once key domains were identified, public international data sources were searched for relevant structured indices to generate a quantitative counterpart (Figure 2).

**Figure 2.** A visual representation of our approach to developing our conceptual framework, from literature review to refinement. NGO: nongovernmental organization.




### Expert Input and Refinement

Our conceptual framework, key domains, and indices were shared formally and informally through email, conversations, and presentations with international colleagues in health care, information technology, informatics, health policy, and international relations. Feedback was collected and an iterative approach was used to refine the model and the indices included.

## Results

### Key Domains and Indices

Our analysis resulted in an inventory of concepts that aggregated into 5 key domains that make up our conceptual framework for assessing health data security considerations in global health partnerships:

1. State of HIT
2. Economics of Health Care
3. Demographics and Equity
4. Societal Freedom and Safety
5. Partnership and Trust

For each of the first 4 domains, we identified publicly available indices that could be used to provide qualitative or quantitative insights. For the fifth domain, *Partnership and Trust*, we created a self-assessment tool that would allow an organization's members to take into consideration the specifics of their project, their experience on the ground, and their in-country partners.

### State of HIT

#### *Overview*

HIT refers to all the electronic systems used by care providers, public health workers, patients, researchers, and others to manage health information [15]. It includes EHRs, other software platforms, hardware devices, and technology systems [16]. HIT systems are a critical component of several aspects of health care, including care delivery, billing, medicolegal liability, research, and health policy [17,18]. HIT use is influenced by factors such as cost, technical feasibility, and regulation [3].

The *State of HIT* domain evaluates a health system's maturity in terms of its logistic, technological, and regulatory progress within a country. The legal structure and regulatory framework around HIT and medical ethics directly influence the viability and progress of HIT [19]. As suggested by Luna et al [19], overcoming legal and ethical challenges, interoperability issues, and technical security vulnerabilities greatly affects the implementation of HIT.

#### *Rationale for Indices*

Multimedia Appendix 1 [6] presents the proposed indices of this domain categorized into 3 broad subdomains:

1. EHR deployment considers the presence of a national EHR system's existence, regulation, and location. National EHR systems require a high level of commitment as well as technical and financial resources [3,20]. This implies a certain level of investment, commitment, and organizational sophistication and provides insight into the groundwork



XSL•FO
RenderX

laid by the country that affects the long-term viability and integrity of HIT and health data [20,21]. Were at al [20] have commented on the unsustainability of expensive implementations and the need for sustainability to be built into EHR implementations, underscoring the demand that EHRs place on technical and financial resources. This subdomain also considers EHR implementation location (in primary, secondary, or tertiary facilities), which gives an understanding of the level of penetration of the technology in the market. Overall, EHRs are often the first, most expansive, and most expensive component of HIT; as such, they serve as reasonable proxies for HIT overall, including technical and logistical risks with implementing new technologies, where other measures are not available. The indices in Multimedia Appendix 1 [6] were gathered from the Electronic Health Record section of the Atlas of eHealth Country Profiles compiled by the WHO as "Yes, No, Not applicable [6]."

2. eHealth Foundations refers to the national strategy, policies, and funding information for eHealth. These factors are critical to determining the success or failure of HIT and can provide valuable insights into risk stratification of a given environment [22]. These indices were gathered from then eHealth Foundations section of the Atlas of eHealth Country Profiles compiled by the WHO as "Yes, No, Not applicable [6]."
3. Legal Frameworks for eHealth refers to the policies, regulation, and jurisdiction governing the use, quality, sharing, and ownership of health data. Medicolegal concerns are universal, complex challenges of HIT that affect use and the experience of patients, providers, and other stakeholders [20,23]. The existence of legal solutions or mechanisms to address these concerns can provide insight into the preparedness of a country to mitigate HIT and data breaches as well as suggest the recourse that individuals have should a breach occur. For example, Palabindala et al [23] mention that Health Insurance Portability and Accountability Act compliance requires substantial legal, technical, and logistical efforts that ensure the establishment of appropriate measures for unfortunate events. These indices were gathered from the Legal Frameworks for eHealth section of the Atlas of eHealth Country Profiles compiled by the WHO as "Yes, No, Not applicable [6]."

## Economics of Health Care

### *Overview*

The *Economics of Health Care* domain aims to quantify a country's overall investment in health care. This domain evaluates a country's financial and resource investment in health care access, delivery, services, and technology. Government investment directly affects infrastructure, quality of care, and affordability for patients and providers, which in turn significantly influences the successful and sustainable implementation of HIT, especially in LMICs [19,24,25]. Without proper allocation of funding, resources, care standards, and cost-effectiveness for stakeholders, there is a greater chance of failure in terms of long-term success and scalability [20]. As discussed by Luna et al [19], financial and technical sustainability is an important element of HIT integration and therefore needs to be addressed before implementation.

### *Rationale for Indices*

This domain can be further evaluated as 4 subdomains; the individual indices have been presented in Multimedia Appendix 2 [26-43]:

1. Health Care Expenditure considers gross domestic product (GDP), contextualized GDP for health care, and health care expenditure. These indices were gathered from data by the World Bank on GDP, GDP per capita, GDP growth percentage, health care spending percentage per GDP, and health care spending per capita in US dollars, US dollars per capita, or percentages depending on the metrics involved [26-31]. Indices such as recontextualized GDP values and health care spending per capita have been described by Naik et al [44] as macroeconomic determinants of health that ultimately influence HIT integration and management. Insufficient health care funding increases the chances of failure of HIT implementation and health data management [20]. Health expenditure and GDP spending on health care thus become potential proxies for risk assessment.
2. Health Care Structure considers both infrastructure and system indices such as hospital beds, doctors per capita, health care access, and the existence of a public health care system. These indices were gathered from data sets by the World Bank on doctors per capita and hospital beds, the Global Burden of Disease (GBD) index on health care access and quality, and the US Social Security Administration on private versus public health care systems as numbers (per 1000 people), indices (1-100), and binary values [32-35]. As these indices directly measure access and the availability of crucial health care resources, they provide insight into the economic and material context of health care. Lower scores for these indices may highlight a higher risk of not having adequate finances and resources for implementing and managing HIT systems [3]. Specifically, Akhlaq et al [3] identified infrastructure, finance, organization, and data management as key factors in the adoption and management of HIT.
3. Health Care Cost considers out-of-pocket fees and universal health coverage. These indices were gathered from the 2017 Global Monitoring Report by the World Bank on universal health coverage as values (0-100) measuring affordability and data sets by the World Bank on out-of-pocket costs as a percentage of total universal coverage [36,37]. These indices are important because they provide insight into the patient-level microeconomic context. Cost-effectiveness and financial viability for patients directly affect access, use, and long-term potential of health care services and resources, including HIT [19]. A lack of affordable health care can create risks for the overall success of health HIT.
4. Health Care Quality is intended to evaluate overall health care system performance through process and outcome measures such as health performance index, infant and maternal mortality, life expectancy, immunization rates, and diarrheal disease rates. These indices were gathered from data sets by GBD collaborators, the WHO, the Central Intelligence Agency, and the GBD database as an index



XSL•FO
RenderX

from 1 to 191, deaths per births, age of death, and percentage immunized [38-43]. The indices listed in Multimedia Appendix 2 [26-43] are frequently used in the literature and by international organizations to measure overall health care quality [25,45,46]. Measuring health care quality provides a lens through which to interpret the economic inputs of a health care system. Major discrepancies between economic inputs and health care quality outcomes may be a cause for concern because these may result from a variety of issues, including health care administration, system stability, and inadequate data collection and reporting. All of these would be factors that may affect health data security.

## Demographics and Equity

### *Overview*

The *Demographics and Equity* domain aims to describe the relevant population and possible disparities involving health care. Understanding how patients from different social, economic, ethnic, and cultural backgrounds experience HIT is important for any global health data partnership because these differences can drive care disparities [25,47]. Gathering together concepts of population demographics, social structure, and community development provides a starting point to gain necessary insights. These social and demographic variables help contextualize pragmatic concerns surrounding patient privacy, access, health discrepancies, and improper use of health data [20]. Increased digitization of health care has several advantages such as public health surveillance during COVID-19, but this same surveillance infrastructure has implications for civil liberties and governance that affect marginalized groups differently [48]. For example, the Social Science Research Council states that Black and Brown communities are subject to disproportionate police surveillance and may be unable to opt out of medical tracking and monitoring systems [48]. Being aware of these issues in the local context is an important component of responsible data stewardship.

### *Rationale for Indices*

Multimedia Appendix 3 [49-61] presents the indices of this domain categorized into 3 broad subdomains:

1. Population metrics include information about the density and structure of the population. The indices were gathered from data by the World Bank on population age structure as percentages and from the United Nations on rural and urban population density in thousands [49-52]. These population metrics affect HIT in a variety of ways. For example, countries with larger populations may require more costly efforts to ameliorate data misuse [62,63]. In addition, differences in age structure and trends may affect the demand and risk of health services and data technology [64-66]. As noted by Knickman and Snell [64], the Baby Boom generation is expected to double by 2030 and will require substantially more health-related resources. The increasing financial demand and use of health data may affect the technical and logistical risks associated with HIT.
2. Social Structure includes information about wealth inequality, poverty, decentralization, and public trust. The indices were gathered from data by the World Bank, a policy paper by World Bank affiliates titled How Close Is Your Government to its People, and Edelman, a global communications firm, as either a percentage or score (1-100), as appropriate [53-56]. Wealth inequality and poverty data provide insights into the economic aspect of inequity. Decentralization has long been advocated by international development agencies to improve health system performance in LMICs [67]. A recent literature review showed limited empirical data to support this approach, but as a dominant theory in international development, it should still be considered [68]. A lack of public trust in government and health care systems can lead to poor patient compliance with public health guidance, delaying seeking care, and withholding of critical information from providers [69]. This can lead to incomplete or unreliable data. Low public trust can be a symptom of either systematic failures of health systems or breaches of trust at the interpersonal level, both of which should be taken into consideration when discussing data privacy and security.
3. Community Development includes information on human development, access to electricity and the internet, literacy rate, and social media penetration. The Human Development Index is a composite measure developed by the United Nations that quantifies the capability of an individual to live a long and healthy life and acquire resources for a basic standard of living as a value from 0 to 1 [57]. Other descriptors of community development that focus on the ability of a community to meaningfully leverage technology are also included. Data sets from the World Bank on internet subscribers, access to electricity, and literacy were gathered as a number (1-100) or percentage as appropriate [58-60]. Social media penetration was gathered from Statcounter Global Stats as a percentage [61]. Basic resources such as electricity and the internet are necessary to meaningfully interact with HIT. This is true at the level of health care facilities as well as at the individual level [2]. Settings with limited access to the internet and electricity may not be able to implement a wide variety of privacy and security tools such as two-factor authentication. Limited literacy can be a barrier to people's ability to use and access technology and data, which may make them more vulnerable to exploitation [70,71]. This increases the human cost of inadequate privacy and security in health care. More specific concepts such as health and technology literacy may be relevant and should be explored further [70,71].

## Societal Freedom and Safety

### *Overview*

Societal freedom is the liberty of an individual to function in society without coercion; the Cato Institute defines this as "the dignity of an individual [72]." As an ever-present societal factor, it influences aspects of health care both directly and indirectly [20,24,47,73]. The *Societal Freedom and Safety* domain aims to quantify the absence of coercive societal constraints on individual freedom within a country as well as the robustness of civil and political liberties; it includes concepts such as liberty of expression, social organization, and lawfulness. Overall,



XSL•FO
RenderX

countries with more freedom (democratic) have more robust health care systems and lower mortality than countries with less freedom (autocratic) [74,75]. Factors such as the rule of law and the influence of civil society affect health and health outcomes [76-78]. Pinzon-Rondon et al [76] found that adherence to the rule of law is associated with a healthier population, higher life expectancy, and lower adverse health outcomes. These social parameters provide insights into how likely a malicious data breach might be; whether the threat to personal sensitive data might come from government or nongovernment actors; and how vulnerable individuals may be to adverse social, financial, and legal consequences in case of a breach of privacy with respect to their personal health data.

### *Rationale for Indices*

Multimedia Appendix 4 [72,79] presents the indices of this domain along with an overall rank and score:

1. Personal Freedom includes information on the rule of law, safety and security, religious freedom, assembly and association, expression, and identity and relationship. These indices were gathered from the Human Freedom Index 2019 by the Cato Institute as an index value between 0 and 1 [72]. Personal freedom is important to consider for data security because it highlights societal challenges with data. For example, religious hostility and persecution, surveillance of expression and information, geopolitical concerns, and stigma toward historically marginalized groups may increase the risk associated with access to identifiable health information [7,80]. In a 2010 report by the London School of Economics, the authors stressed the importance of social context and appropriate safeguards for HIT implementation, given the vast cultural and environmental differences that can exist even within a country [80]. In addition, limitations in assembly and association may adversely influence public health initiatives and health policy [78]. These concepts can provide a more nuanced assessment of the data security risk.
2. Economic Freedom includes economic liberty, sound money, property rights, international trade, and regulation of financial institutions. These indices have been gathered from the Human Freedom Index 2019 by the Cato Institute as an index value between 0 and 1 [72]. Economic freedom provides insight into the financial opportunities for individuals and organizations. In societies with high economic freedom there may be additional economic incentives to develop robust HIT [81]. There may also be opportunities for private and public-private partnerships to enhance data security [82].
3. Global Freedom is a concept developed by Freedom House, a US-based nonprofit focused on promoting democracy, and published in their annual Freedom in the World Report since 1973 [79,83]. It is a quantitative and qualitative assessment of political rights and civil liberties in countries and territories around the world, represented as a weight score on a scale from –4 to 100. Evaluations of a country's electoral process, political participation, government functioning, associational rights, rule of law, and personal autonomy make up the global freedom score. Freedom House also publishes other relevant indices such as the Internet Freedom Score and Democracy Score, but these cover a significantly smaller number of the world's countries (65 countries and 29 countries, respectively) and thus may not be as helpful in creating a standard analytical approach.

## Partnership and Trust

An increasing number of partnerships have been developing between high-income countries and LMICs to address the global burden of disease. The success of these projects requires strong partnerships, which involves establishing rapport [84]. Relationship building and trust have been shown to be critical in navigating pragmatic obstacles as well as cultural and logistical boundaries [85,86]. Wagner et al [85] highlight how local coordinators and hosting communities are vital for the execution of international projects, and therefore establishing relationships and promoting ongoing collaboration are imperative to the success of global health efforts.

The strength of partnerships may be a moderating factor for concerns around patient data misuse. Organizations should objectively evaluate their global health partnerships. However, there is limited literature on global health partnership assessment tools. Instruments such as the Partnership Assessment Toolkit do exist, but studies need to be conducted to better understand their uses, limitations, and effectiveness [87,88]. To our knowledge, there are no tools that specifically address health data concerns. Given that each partnership is unique and influenced by a number of factors, we propose that this domain should be a self-assessment completed by the collaborating organizations. Relevant questions to explore include details about the in-country partner, the length of time the partnership has existed, the scope of the partnership, sensitive personal data collection, and relevant data security expertise resources available to the partners. Textbox 2 presents a list of potential questions to include in a self-assessment.



XSL•FO
RenderX

**Textbox 2.** Potential questions for Partnership and Trust self-assessment.

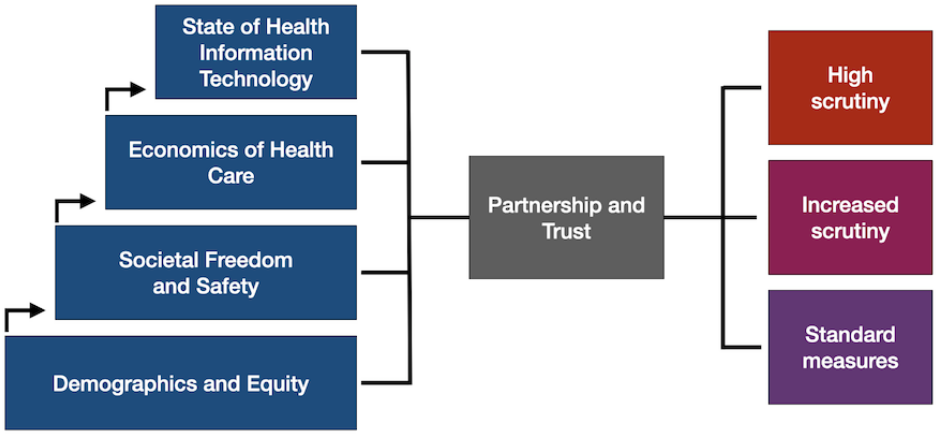
**Questions to include in a self-assessment**

- How long have your organizations worked together?
- How long has your partner been active in-country?
- How much experience does your partner have with health data?
- Is your local partner in good standing in-country?
- What is your organization's reputation in-country?
- Is an official government entity with oversight over health, health care, or data involved in your project? If not, should they be?
- Are there known examples of health data misuse in the country?
- In your partnership, who owns the data?
- Who is responsible for data security?
- What, if any, sensitive patient data are being collected or used?
- What physical, technical, and procedural measures have been taken to safeguard patient data?
- What data safety and monitoring measures will be put in place?
- Do you and your partner have the relevant experience to serve as data stewards?

### Conceptual Framework

Figure 3 shows the relationship among these 5 domains and how they might be leveraged to provide a risk stratification of global health data partnerships. This type of conceptual framework has been leveraged to address various issues in informatics, such as the development of global health networks, patient safety, and conceptual models for research [89-91]. The value of these conceptual frameworks is in laying out all the components of a given issue, exploring their interrelationships, and identifying the emerging complexity [92]. In the current framework, *Demographics and Equity* and *Societal Freedom* provide a foundational understanding of a given country. *Economics of Health Care* can be understood within that context, and the *State of HIT* is influenced by, and builds upon, all 3 domains. *Partnership and Trust* is a moderating factor for all other variables. A long-standing, effective partnership with high levels of trust and cooperation may overcome a number of deficiencies in other domains, whereas an unstable or ineffective partnership may suffer from serious data concerns despite an otherwise favorable environment. The latter case is often the cause for HIT implementation failures in high-income countries [93].

**Figure 3.** Four of the domains build on each other (Demographics and Equity, Economics of Health Care, Societal Freedom and Safety, and State of Health Information Technology). These are moderated by the fifth domain, Partnership and Trust. Together, these domains can be evaluated to produce a risk stratification for global health data partnerships and health information technology projects.




## Discussion

### Overview

As HIT deployments continue to progress in LMICs, data security concerns will become more prevalent. The development of this conceptual framework is an attempt to better understand the many variables that might affect health data security in a given country. There are a number of existing models for assessing the maturity of HIT and data security [94]. However, most of these models have been developed or applied in high-income countries and make assumptions about the legal, regulatory, and technical capacity already in place; these assumptions often do not hold true in LMICs. The health systems of high-income countries (and the social, political, and economic forces that support them) can vary significantly from



XSL•FO
RenderX

those in LMICs. For example, in the review by Tarhan et al [94], the authors provide the full list of countries in which the maturity models they reviewed were developed and applied. The average WHO health performance index of these countries was 0.813 (SD 0.17), whereas the average health performance index of CIS countries is 0.612 (SD 0.12; $P$=.001 by 2-tailed, 2 sample $t$ test) [38]. Therefore, it may not be meaningful to apply existing maturity models to LMICs; to our knowledge, there are no models explicitly developed for low-resource settings.

The goal of our research is ultimately to develop an assessment and decision support tool that organizations can use in their global health partnerships. In the first version of this tool, we anticipate a more qualitative approach in which organizations use these domains to guide them in conducting a thorough evaluation of projects and partnerships. A second version of the tool will have a more quantitative component; 4 of the 5 key domains we identified use publicly available indices that could be integrated into a mathematical model that describes health data risk. Multimedia Appendix 5 presents the values for a sampling of indices across the domains of *State of HIT*, *Economics of Health Care*, *Demographics and Equity*, and *Societal Freedom*. We have included 9 CIS countries for which data are available (Armenia, Azerbaijan, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, and Ukraine). In addition, data have been aggregated for several other countries to serve as reference points for the 4 domains of interest. These range from countries that have a track record of excellence in public health (Germany, France, Switzerland, South Korea, and Japan) to countries that have made significant advances in HIT (Estonia) and modernized their health care systems in other geographic regions (China and South Africa), as well as the United States. For the fifth domain, *Partnership and Trust*, additional research and validation will be required to finalize the self-assessment questionnaire.

The concept of bias was a recurring theme in many of the sources we evaluated. During the COVID-19 epidemic, the disproportionate impact that bias, discrimination, and racism can have on health outcomes was highlighted at scale [95]. These social biases can easily be translated into HIT; data sets, algorithms, and predictive models are all subject to both explicit and implicit bias [70]. The types of individuals who are included in data sets, the data elements that are and are not collected, the variables that are highlighted, and the outcomes that are selected for can all result in HIT applications that adversely affect care delivery and further drive health disparities among marginalized groups [70,96,97]. Although not directly linked to data security considerations (except to the extent to which security failures may adversely affect members of marginalized groups, as discussed previously in other domains), we believe that an awareness of bias is critical in any global health data partnership. Additional research is needed to identify approaches to measure and account for biases that may differ across settings.

## Limitations

Our study includes several limitations. Although we reviewed the literature to inform our research and approach, there is still a need for a comprehensive systematic literature review to be conducted and published. Given the evolving nature of the subject, a scoping review methodology would be appropriate; our research team is preparing to embark on this project [98]. Our proposed domains reflect our research and experience but need further validation from the broader community of health care, HIT, and public policy professionals. Our conceptual framework has not been tested qualitatively or quantitatively against real-world examples; thus, it remains to be seen if it can meaningfully capture the complexities and nuances of health data security. Finally, it is unclear if our proposed indices will result in a useful quantitative model of risk; further analysis is required.

## Next Steps

To advance our research agenda, we plan to engage in the following activities:

1. Validate our conceptual framework: Additional work is needed to validate our proposed framework. We will share our framework publicly to gather both formal and informal feedback from stakeholders around the world. We will also begin the work of applying the framework to real-world examples in collaboration with local experts to test its internal and external validity. Further literature review and qualitative research will be needed to finalize the Partnership and Trust self-assessment.
2. Develop a qualitative assessment tool: Once we have finalized and validated the framework, we will develop a qualitative assessment tool that other organizations can use to evaluate their existing data partnerships. This stage will not only provide additional external validation and refinement of the framework, but will also provide the opportunity to conduct user-centered design research to improve the usability of the tool and related documentation.
3. Develop a quantitative model: We plan to work with our data science colleagues to use both traditional statistical methods and more modern machine learning approaches to develop a quantitative model of our conceptual framework. This will require extensive validation, but if successful, it may result in a risk stratification that could conceivably be calculated for every country, needing only the Partnership and Trust self-assessment to provide local context.
4. Dissemination of findings: Our overall goal is to support how organizations make decisions around global health data partnerships. The current framework, assessment tools, and quantitative models are all intended to help organizations make the best decisions possible in terms of safeguarding patient data in LMICs. To that end, we intend to use multiple avenues to disseminate our work, including publications, presentations, webinars, and white papers. We plan to collaborate with universities and nongovernmental organizations to help them to implement the conceptual framework and associated tools.

## Conclusions

Global health HIT partnerships have the potential to have a positive impact in LMICs, leveraging the resources and knowledge of partner organizations to build in-country capacity and expertise. However, there are gaps in the legal, technical,



XSL•FO
RenderX

and regulatory environments in many LMICs, increasing the risk of possible health data misuse, particularly among marginalized and vulnerable populations. Our conceptual framework helps to identify key domains that may have an impact on health data security considerations in global health partnerships. Additional research is needed to further validate and improve the framework. We encourage global health, HIT, and health care professionals to participate in improving this framework. In the future, we hope to be able to leverage our framework to create assessment and decision-making tools that can be used to evaluate risk in other global health initiatives such as clinical and academic partnerships, pandemic control, and emergency response operations.

## Acknowledgments

The authors would like to acknowledge Mr Pushkar Sharma for his insights into quantitative indices of freedom and democracy around the world, Dr Laura Ferguson for her input on quality-of-care standards, and Dr Neeraj Sood and Mr Virat Agrawal for their help in refining the overall concepts presented and helping to identify additional quantitative indices. The authors would also like to express their gratitude to all the members of the Online Pediatric Education Network and Avetis team at Children's Hospital Los Angeles for their camaraderie and support.

## Authors' Contributions

JE conceived of the overall study, outlined the manuscript, contributed to the writing and editing of the manuscript, and oversaw the project. ATS generated the first draft, performed the initial literature review, and assembled the textboxes. JD and TL contributed to the editing of the manuscript and provided feedback and subject matter expertise.

## Conflicts of Interest

JE is a subject matter consultant for AI Health, a machine learning company that focuses on chronic disease management. JD holds equity in a health data company, Orderly Health. JD also serves as an advisor for an electronic medical record start up, Frontida Health.

Multimedia Appendix 1
State of health information technology domain.
[PDF File (Adobe PDF File), 17 KB - formative_v5i12e25833_app1.pdf ]

Multimedia Appendix 2
Economics of healthcare domain.
[PDF File (Adobe PDF File), 34 KB - formative_v5i12e25833_app2.pdf ]

Multimedia Appendix 3
Demographics and equity domain.
[PDF File (Adobe PDF File), 30 KB - formative_v5i12e25833_app3.pdf ]

Multimedia Appendix 4
Societal freedom and safety domain.
[PDF File (Adobe PDF File), 29 KB - formative_v5i12e25833_app4.pdf ]

Multimedia Appendix 5
Example index values for CIS and reference countries.
[PDF File (Adobe PDF File), 20 KB - formative_v5i12e25833_app5.pdf ]

## References

1. Electronic health records. World Health Organization. 2021. URL: https://www.who.int/gho/goe/electronic_health_records/en/ [accessed 2021-01-21]
2. Syzdykova A, Malta A, Zolfo M, Diro E, Oliveira JL. Open-source electronic health record systems for low-resource settings: systematic review. JMIR Med Inform 2017 Nov 13;5(4):e44 [FREE Full text] [doi: 10.2196/medinform.8131] [Medline: 29133283]
3. Akhlaq A, McKinstry B, Muhammad KB, Sheikh A. Barriers and facilitators to health information exchange in low- and middle-income country settings: a systematic review. Health Policy Plan 2016 Dec;31(9):1310-1325. [doi: 10.1093/heapol/czw056] [Medline: 27185528]



XSL•FO
RenderX

4. Tierney W, Achieng M, Baker E, Bell A, Biondich P, Braitstein P, et al. Experience implementing electronic health records in three East African countries. Stud Health Technol Inform (PART 1) 2010;160:375. [doi: 10.3233/978-1-60750-588-4-371]
5. Jawhari B, Ludwick D, Keenan L, Zakus D, Hayward R. Benefits and challenges of EMR implementations in low resource settings: a state-of-the-art review. BMC Med Inform Decis Mak 2016 Sep 06;16:116 [FREE Full text] [doi: 10.1186/s12911-016-0354-8] [Medline: 27600269]
6. WHO Global Observatory for eHealth. Atlas of eHealth Country Profiles: The Use of eHealth in Support of Universal Health Coverage. Geneva: World Health Organization; 2016:1-392.
7. Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi L. Big data in global health: improving health in low- and middle-income countries. Bull World Health Organ 2015 Jan 30;93(3):203-208. [doi: 10.2471/blt.14.139022]
8. CIS Executive Committee. Executive Committee of the Commonwealth of Independent States. 2021. URL: https://cis.minsk.by/ [accessed 2021-01-21]
9. Rechel B, Roberts B, Richardson E, Shishkin S, Shkolnikov VM, Leon DA, et al. Health and health systems in the Commonwealth of Independent States. Lancet 2013 Mar;381(9872):1145-1155. [doi: 10.1016/s0140-6736(12)62084-4]
10. Sheiman I, Shishkin S, Shevsky V. The evolving Semashko model of primary health care: the case of the Russian Federation. Risk Manag Healthc Policy 2018 Nov;Volume 11:209-220. [doi: 10.2147/rmhp.s168399]
11. Sheiman I. Rocky road from the Semashko to a new health model. Interview by Fiona Fleck. Bull World Health Organ 2013 May 01;91(5):320-321 [FREE Full text] [doi: 10.2471/BLT.13.030513] [Medline: 23678194]
12. Rechel B, Richardson E, McKee M. Trends in health systems in the former Soviet countries. Eur J Public Health 2014;24(suppl_2):1-21. [doi: 10.1093/eurpub/cku162.088]
13. Jabareen Y. Building a conceptual framework: philosophy, definitions, and procedure. Int J Qual Methods 2009 Dec 01;8(4):49-62. [doi: 10.1177/160940690900800406]
14. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
15. Health information technology. U.S. Department of Health & Human Services. 2020. URL: https://www.hhs.gov/hipaa/for-professionals/special-topics/health-information-technology/index.html [accessed 2021-01-21]
16. Health IT: advancing America's health care. The Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/sites/default/files/pdf/health-information-technology-fact-sheet.pdf [accessed 2021-01-21]
17. Yen P, McAlearney AS, Sieck CJ, Hefner JL, Huerta TR. Health Information Technology (HIT) adaptation: refocusing on the journey to successful HIT implementation. JMIR Med Inform 2017 Sep 07;5(3):e28 [FREE Full text] [doi: 10.2196/medinform.7476] [Medline: 28882812]
18. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2018 Mar 06;25(S 01):48-61. [doi: 10.15265/iys-2016-s006]
19. Luna D, Almerares A, Mayan JC, González BD, Otero C. Health informatics in developing countries: going beyond pilot practices to sustainable implementations: a review of the current challenges. Healthc Inform Res 2014 Jan;20(1):3-10 [FREE Full text] [doi: 10.4258/hir.2014.20.1.3] [Medline: 24627813]
20. Were MC, Meslin EM. Ethics of implementing Electronic Health Records in developing countries: points to consider. AMIA Annu Symp Proc 2011;2011:1499-1505 [FREE Full text] [Medline: 22195214]
21. Boonstra A, Versluis A, Vos JF. Implementing electronic health records in hospitals: a systematic literature review. BMC Health Serv Res 2014 Sep 04;14(1):370 [FREE Full text] [doi: 10.1186/1472-6963-14-370] [Medline: 25190184]
22. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of ehealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
23. Palabindala V, Pamarthy A, Jonnalagadda NR. Adoption of electronic health records and barriers. J Community Hosp Intern Med Perspect 2016 Oct 26;6(5):32643 [FREE Full text] [doi: 10.3402/jchimp.v6.32643] [Medline: 27802857]
24. Ozair FF, Jamshed N, Sharma A, Aggarwal P. Ethical issues in electronic health records: a general overview. Perspect Clin Res 2015;6(2):73-76 [FREE Full text] [doi: 10.4103/2229-3485.153997] [Medline: 25878950]
25. Carney TJ, Kong AY. Leveraging health informatics to foster a smart systems response to health disparities and health equity challenges. J Biomed Inform 2017 Apr;68:184-189 [FREE Full text] [doi: 10.1016/j.jbi.2017.02.011] [Medline: 28214562]
26. Population data. United Nations. URL: https://population.un.org/wpp/Download/Standard/Population/ [accessed 2021-01-21]
27. GDP (current US$). The World Bank. 2021. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD [accessed 2021-01-21]
28. GDP per capita (current US$). The World Bank. 2021. URL: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD [accessed 2021-01-21]
29. GDP growth (annual %). The World Bank. 2021. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG [accessed 2021-01-21]
30. Current health expenditure (% of GDP). The World Bank. 2021. URL: https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS [accessed 2021-01-21]




XSL•FO
RenderX

31. Current health expenditure per capita (current US$). The World Bank. 2021. URL: https://data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD [accessed 2021-01-21]
32. Physicians (per 1,000 people). The World Bank. 2021. URL: https://data.worldbank.org/indicator/SH.MED.PHYS.ZS [accessed 2021-01-21]
33. Hospital beds (per 1,000 people). The World Bank. 2021. URL: https://data.worldbank.org/indicator/SH.MED.BEDS.ZS [accessed 2021-01-21]
34. Global Burden of Disease Study 2015 (GBD 2015) healthcare access and quality index based on amenable mortality 1990–2015. Global Burden of Disease Collaborative Network. 2019. URL: http://ghdx.healthdata.org/record/ihme-data/gbd-2015-healthcare-access-and-quality-index-1990-2015 [accessed 2021-01-21]
35. Country profiles. International Social Security Association (ISSA). URL: https://ww1.issa.int/country-profiles [accessed 2021-01-21]
36. Tracking universal health coverage: 2017 Global Monitoring Report. In: World Health Organization and the International Bank for Reconstruction and Development / The World Bank 2017. Geneva: World Health Organization and The World Bank; 2017.
37. Out-of-pocket expenditure (% of current health expenditure). The World Bank. 2021. URL: https://data.worldbank.org/indicator/SH.XPD.OOPC.CH.ZS [accessed 2021-01-21]
38. Tandon A, Murray C, Lauer J, Evans D. Measuring overall health system performance for 191 countries. EIP/GPE/EQC - World Health Organization. URL: https://www.who.int/healthinfo/paper30.pdf [accessed 2021-01-21]
39. Infant mortality rate - the world factbook. Central Intelligence Agency. URL: https://www.cia.gov/the-world-factbook/field/infant-mortality-rate/country-comparison [accessed 2021-01-21]
40. Maternal mortality ratio (per 100 000 live births). World Health Organization. 2021. URL: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/maternal-mortality-ratio-(per-100-000-live-births) [accessed 2021-01-21]
41. Life expectancy and healthy life expectancy. World Health Organization. 2021. URL: https://www.who.int/data/gho/data/themes/topics/indicator-groups/indicator-group-details/GHO/life-expectancy-and-healthy-life-expectancy [accessed 2021-01-21]
42. WHO vaccine-preventable diseases: monitoring system. 2020 global summary. World Health Organization. 2020. URL: https://apps.who.int/immunization_monitoring/globalsummary [accessed 2021-01-21]
43. GBD results tool. Global Health Data Exchange. 2021. URL: http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/825168e248c37b282cc57a0e9ce957e3 [accessed 2021-01-21]
44. Naik Y, Baker P, Walker I, Tillmann T, Bash K, Quantz D, et al. The macro-economic determinants of health and health inequalities-umbrella review protocol. Syst Rev 2017 Nov 03;6(1):222 [FREE Full text] [doi: 10.1186/s13643-017-0616-2] [Medline: 29100497]
45. Kruk M, Gage A, Joseph N, Danaei G, García-Saisó S, Salomon J. Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. Lancet 2018 Nov;392(10160):2203-2212. [doi: 10.1016/s0140-6736(18)31668-4]
46. Rémy V, Zöllner Y, Heckmann U. Vaccination: the cornerstone of an efficient healthcare system. J Mark Access Health Policy 2015;3:10 [FREE Full text] [doi: 10.3402/jmahp.v3.27041] [Medline: 27123189]
47. Sulmasy LS, López AM, Horwitch CA, , American College of Physicians Ethics, ProfessionalismHuman Rights Committee. Ethical implications of the electronic health record: in the service of the patient. J Gen Intern Med 2017 Aug;32(8):935-939 [FREE Full text] [doi: 10.1007/s11606-017-4030-1] [Medline: 28321550]
48. Surveillance and the 'New Normal' of Covid-19: public health, data, and justice. Social Science Research Council. 2021. URL: https://covid19research.ssrc.org/public-health-surveillance-and-human-rights-network/report/ [accessed 2021-01-21]
49. Population ages 0-14 (% of total population). The World Bank. 2019. URL: https://data.worldbank.org/indicator/SP.POP.0014.TO.ZS [accessed 2021-01-21]
50. Population ages 15-64 (% of total population). The World Bank. 2019. URL: https://data.worldbank.org/indicator/SP.POP.1564.TO.ZS [accessed 2021-01-21]
51. Population ages 65 and above (% of total population). The World Bank. 2019. URL: https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS [accessed 2021-01-21]
52. World Urbanization Prospects: the 2018 revision. United Nations, Department of Economic and Social Affairs, Population Division. 2018. URL: https://population.un.org/wup/Download/ [accessed 2021-01-21]
53. Gini index (World Bank estimate). The World Bank. 2021. URL: https://data.worldbank.org/indicator/SI.POV.GINI [accessed 2021-01-21]
54. Poverty headcount ratio at national poverty lines (% of population). The World Bank. 2021. URL: https://data.worldbank.org/indicator/SI.POV.NAHC [accessed 2021-01-21]
55. Ivanyna M, Shah A. How close is your government to its people? Worldwide indicators on localization and decentralization. Economics 2014;8(1):1-61. [doi: 10.5018/economics-ejournal.ja.2014-3]
56. Edelman trust barometer. Edelman. 2020. URL: https://www.edelman.com/sites/g/files/aatuss191/files/2020-01/2020%20Edelman%20Trust%20Barometer%20Global%20Report_LIVE.pdf [accessed 2021-01-21]
57. Human Development Report 2019. New York: United Nations Development Programme; 2019.




XSL•FO
RenderX

58. Fixed broadband subscriptions (per 100 people). The World Bank. 2021. URL: https://data.worldbank.org/indicator/IT.NET.BBND.P2 [accessed 2021-01-21]
59. Access to electricity (% of population). The World Bank. 2021. URL: https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS [accessed 2021-01-21]
60. Literacy rate, adult total (% of people ages 15 and above). The World Bank. 2020. URL: https://data.worldbank.org/indicator/SE.ADT.LITR.ZS [accessed 2021-01-20]
61. Social media stats worldwide. StatCounter Global Stats. 2021. URL: https://gs.statcounter.com/social-media-stats [accessed 2021-01-21]
62. Seh AH, Zarour M, Alenezi M, Sarkar AK, Agrawal A, Kumar R, et al. Healthcare data breaches: insights and implications. Healthcare (Basel) 2020 May 13;8(2):133 [FREE Full text] [doi: 10.3390/healthcare8020133] [Medline: 32414183]
63. Cost of a data breach report. IBM Security. 2020. URL: https://www.capita.com/sites/g/files/nginej146/files/2020-08/Ponemon-Global-Cost-of-Data-Breach-Study-2020.pdf [accessed 2021-01-21]
64. Knickman J, Snell E. The 2030 problem: caring for aging baby boomers. Health Serv Res 2002 Aug;37(4):849-884 [FREE Full text] [doi: 10.1034/j.1600-0560.2002.56.x] [Medline: 12236388]
65. Canizares M, Gignac M, Hogg-Johnson S, Glazier RH, Badley EM. Do baby boomers use more healthcare services than other generations? Longitudinal trajectories of physician service use across five birth cohorts. BMJ Open 2016 Sep 29;6(9):e013276 [FREE Full text] [doi: 10.1136/bmjopen-2016-013276] [Medline: 27687902]
66. Villa VM, Wallace SP, Bagdasaryan S, Aranda MP. Hispanic Baby Boomers: health inequities likely to persist in old age. Gerontologist 2012 May;52(2):166-176 [FREE Full text] [doi: 10.1093/geront/gns002] [Medline: 22399578]
67. Faguet J. Decentralization and governance. World Dev 2014 Jan;53:2-13 [FREE Full text] [doi: 10.1016/j.worlddev.2013.01.002]
68. Dwicaksono A, Fox AM. Does decentralization improve health system performance and outcomes in low- and middle-income countries? A systematic review of evidence from quantitative studies. Milbank Q 2018 Jun 04;96(2):323-368 [FREE Full text] [doi: 10.1111/1468-0009.12327] [Medline: 29870116]
69. Gille F, Smith S, Mays N. Why public trust in health care systems matters and deserves greater research attention. J Health Serv Res Policy 2015 Jan;20(1):62-64. [doi: 10.1177/1355819614543161] [Medline: 25038059]
70. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
71. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
72. Vásquez I, Porčnik T. The Human Freedom Index 2019. Washington: Cato Institute; 2019.
73. Carroll A, Quinn S. Forced Out: LGBT People in Armenia - Report on ILGA-Europe/COC Fact-finding Mission. Belgium: ILGA-Europe; 2009:1-82.
74. Bollyky T, Templin T, Cohen M, Schoder D, Dieleman J, Wigley S. The relationships between democratic experience, adult health, and cause-specific mortality in 170 countries between 1980 and 2016: an observational analysis. Lancet 2019 Apr;393(10181):1628-1640. [doi: 10.1016/s0140-6736(19)30235-1]
75. Bollyky TJ. Democracy matters in global health. Council on Foreign Relations. 2019. URL: https://www.cfr.org/article/democracy-matters-global-health [accessed 2021-01-21]
76. Pinzon-Rondon AM, Attaran A, Botero JC, Ruiz-Sternberg AM. Association of rule of law and health outcomes: an ecological study. BMJ Open 2015 Oct 29;5(10):e007004 [FREE Full text] [doi: 10.1136/bmjopen-2014-007004] [Medline: 26515684]
77. Dingake OB. The rule of law as a social determinant of health. Health Hum Rights 2017 Dec;19(2):295-298 [FREE Full text] [Medline: 29302183]
78. Gómez EJ. Civil society in global health policymaking: a critical review. Global Health 2018 Jul 25;14(1):73 [FREE Full text] [doi: 10.1186/s12992-018-0393-2] [Medline: 30045738]
79. Freedom in the world research methodology. Freedom House. 2021. URL: https://freedomhouse.org/reports/freedom-world/freedom-world-research-methodology [accessed 2021-01-21]
80. Electronic health privacy and security in developing countries and humanitarian operations. The London School of Economics and Political Science. URL: https://personal.lse.ac.uk/martinak/ehealth.pdf [accessed 2021-01-21]
81. Adler-Milstein J, Jha AK. HITECH act drove large gains in hospital electronic health record adoption. Health Aff (Millwood) 2017 Aug 01;36(8):1416-1422. [doi: 10.1377/hlthaff.2016.1651] [Medline: 28784734]
82. Gold M, Mc L. Assessing HITECH implementation and lessons: 5 years later. Milbank Q 2016 Sep;94(3):654-687 [FREE Full text] [doi: 10.1111/1468-0009.12214] [Medline: 27620687]
83. Countries and territories - Freedom scores. Freedom House. 2021. URL: https://freedomhouse.org/countries/nations-transit/scores [accessed 2021-01-21]
84. Guilfoyle R, Morzycki AD, Saleh A. What makes global healthcare partnerships successful? A systematic review. Glob Public Health 2021 Mar 09:1-10. [doi: 10.1080/17441692.2021.1892795] [Medline: 33689576]



XSL•FO
RenderX

85. Wagner JP, Schroeder AD, Espinoza JC, Hiatt JR, Mellinger JD, Cusick RA, et al. Global outreach using a systematic, competency-based training paradigm for inguinal hernioplasty. JAMA Surg 2017 Jan 01;152(1):66-73. [doi: 10.1001/jamasurg.2016.3323] [Medline: 27706482]
86. Christopher S, Watts V, McCormick AK, Young S. Building and maintaining trust in a community-based participatory research partnership. Am J Public Health 2008 Aug;98(8):1398-1406. [doi: 10.2105/AJPH.2007.125757] [Medline: 18556605]
87. Frearson A. Partnership self-assessment toolkit: a practical guide to creating and maintaining successful partnerships. East Leeds Primary Care Trust. 2002. URL: https://iapo.org.uk/sites/default/files/current/resources/LHAZ_Partnership_selfassessment_toolkit.pdf [accessed 2021-01-20]
88. Yarmoshuk AN, Guantai AN, Mwangu M, Cole DC, Zarowsky C. What makes International Global Health University partnerships higher-value? An examination of partnership types and activities favoured at four East African universities. Ann Glob Health 2018 Apr 30;84(1):139-150 [FREE Full text] [doi: 10.29024/aogh.20] [Medline: 30873772]
89. Shiffman J, Quissell K, Schmitz HP, Pelletier DL, Smith SL, Berlan D, et al. A framework on the emergence and effectiveness of global health networks. Health Policy Plan 2016 Apr 29;31 Suppl 1(suppl 1):i3-16 [FREE Full text] [doi: 10.1093/heapol/czu046] [Medline: 26318679]
90. Singh H, Sittig DF. Measuring and improving patient safety through health information technology: the Health IT Safety Framework. BMJ Qual Saf 2016 Apr 14;25(4):226-232 [FREE Full text] [doi: 10.1136/bmjqs-2015-004486] [Medline: 26369894]
91. Gray K, Sockolow P. Conceptual models in health informatics research: a literature review and suggestions for development. JMIR Med Inform 2016 Feb 24;4(1):e7 [FREE Full text] [doi: 10.2196/medinform.5021] [Medline: 26912288]
92. Green HE. Use of theoretical and conceptual frameworks in qualitative research. Nurse Res 2014 Jul;21(6):34-38. [doi: 10.7748/nr.21.6.34.e1252] [Medline: 25059086]
93. Wiehe SE, Rosenman MB, Chartash D, Lipscomb ER, Nelson TL, Magee LA, et al. A solutions-based approach to building data-sharing partnerships. EGEMS (Wash DC) 2018 Aug 22;6(1):20 [FREE Full text] [doi: 10.5334/egems.236] [Medline: 30155508]
94. Tarhan AK, Garousi V, Turetken O, Söylemez M, Garossi S. Maturity assessment and maturity models in health care: a multivocal literature review. Digit Health 2020 Apr 01;6:2055207620914772 [FREE Full text] [doi: 10.1177/2055207620914772] [Medline: 32426151]
95. Gravlee C. Systemic racism, chronic health inequities, and COVID-19: a syndemic in the making? Am J Hum Biol 2020 Sep;32(5):e23482 [FREE Full text] [doi: 10.1002/ajhb.23482] [Medline: 32754945]
96. FitzGerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. BMC Med Ethics 2017 Mar 01;18(1):19 [FREE Full text] [doi: 10.1186/s12910-017-0179-8] [Medline: 28249596]
97. Blair IV, Steiner J, Havranek E. Unconscious (implicit) bias and health disparities: where do we go from here? Perm J 2011;15(2):71-78 [FREE Full text] [Medline: 21841929]
98. Lewinter KE, Hudson SM, Kysh L, Lara M, Betz CL, Espinoza J. Reconsidering reviews: the role of scoping reviews in digital medicine and pediatrics. NPJ Digit Med 2020 Dec 10;3(1):158 [FREE Full text] [doi: 10.1038/s41746-020-00368-2] [Medline: 33303889]

## Abbreviations

**CIS:** Commonwealth of Independent States
**EHR:** electronic health record
**GBD:** Global Burden of Disease
**GDP:** gross domestic product
**HIT:** health information technology
**LMIC:** low- and middle-income country
**WHO:** World Health Organization

*Edited by C Lovis; submitted 29.01.21; peer-reviewed by K Fuji, M Salman; comments to author 23.05.21; revised version received 30.06.21; accepted 10.10.21; published 08.12.21.*

*Please cite as:*
*Espinoza J, Sikder AT, Dickhoner J, Lee T*
*Assessing Health Data Security Risks in Global Health Partnerships: Development of a Conceptual Framework*
*JMIR Form Res 2021;5(12):e25833*
*URL: https://formative.jmir.org/2021/12/e25833*
*doi:10.2196/25833*
*PMID:34889752*



XSL•FO
RenderX

©Juan Espinoza, Abu Taher Sikder, James Dickhoner, Thomas Lee. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Test-Retest Reliability of Home-Based Fitness Assessments Using a Mobile App (R Plus Health) in Healthy Adults: Prospective Quantitative Study

I-I Lin[1], MSc; You-Lin Chen[1], MSc; Li-Ling Chuang[2,3,4], PhD

[1]Recovery Plus Inc, Chengdu, China

[2]School of Physical Therapy & Graduate Institute of Rehabilitation Science, College of Medicine, Chang Gung University, Taoyuan, Taiwan

[3]Department of Physical Medicine and Rehabilitation, Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan

[4]Healthy Aging Research Center, Chang Gung University, Taoyuan, Taiwan

**Corresponding Author:**
Li-Ling Chuang, PhD
School of Physical Therapy & Graduate Institute of Rehabilitation Science, College of Medicine, Chang Gung University
No. 259 Wen-hua 1st Rd, Guishan Dist
Taoyuan, 33302
Taiwan
Phone: 886 3 2118800 ext 3177
Fax: 886 3 2118700
Email: lchuang@gap.cgu.edu.tw

## Abstract

**Background:** Poor physical fitness has a negative impact on overall health status. An increasing number of health-related mobile apps have emerged to reduce the burden of medical care and the inconvenience of long-distance travel. However, few studies have been conducted on home-based fitness tests using apps. Insufficient monitoring of physiological signals during fitness assessments have been noted. Therefore, we developed R Plus Health, a digital health app that incorporates all the components of a fitness assessment with concomitant physiological signal monitoring.

**Objective:** The aim of this study is to investigate the test-retest reliability of home-based fitness assessments using the R Plus Health app in healthy adults.

**Methods:** A total of 31 healthy young adults self-executed 2 fitness assessments using the R Plus Health app, with a 2- to 3-day interval between assessments. The fitness assessments included cardiorespiratory endurance, strength, flexibility, mobility, and balance tests. The intraclass correlation coefficient was computed as a measure of the relative reliability of the fitness assessments and determined their consistency. The SE of measurement, smallest real difference at a 90% CI, and Bland–Altman analyses were used to assess agreement, sensitivity to real change, and systematic bias detection, respectively.

**Results:** The relative reliability of the fitness assessments using R Plus Health was moderate to good (intraclass correlation coefficient 0.8-0.99 for raw scores, 0.69-0.99 for converted scores). The SE of measurement and smallest real difference at a 90% CI were 1.44-6.91 and 3.36-16.11, respectively, in all fitness assessments. The 95% CI of the mean difference indicated no significant systematic error between the assessments for the strength and balance tests. The Bland–Altman analyses revealed no significant systematic bias between the assessments for all tests, with a few outliers. The Bland–Altman plots illustrated narrow limits of agreement for upper extremity strength, abdominal strength, and right leg stance tests, indicating good agreement between the 2 assessments.

**Conclusions:** Home-based fitness assessments using the R Plus Health app were reliable and feasible in young, healthy adults. The results of the fitness assessments can offer a comprehensive understanding of general health status and help prescribe safe and suitable exercise training regimens. In future work, the app will be tested in different populations (eg, patients with chronic diseases or users with poor fitness), and the results will be compared with clinical test results.

**Trial Registration:** Chinese Clinical Trial Registry ChiCTR2000030905; http://www.chictr.org.cn/showproj.aspx?proj=50229

***(JMIR Form Res 2021;5(12):e28040)*** doi:10.2196/28040



XSL•FO
RenderX

**KEYWORDS**

mobile health app; reliability; home-based fitness assessments; healthy adults; mobile phone; digital health

# Introduction

## Background

Physical fitness plays an important role in overall health and quality of life and is directly related to physical activity [1]. Regular physical activity confers health benefits, such as increased life expectancy and reduced mortality [2,3]. However, the World Health Organization has reported that >80% of adolescents globally do not engage in sufficient physical activity. The prevalence of insufficient physical activity was 27.5% among adults aged >18 years worldwide [4]. Studies have indicated that physical inactivity is associated with poor physical fitness and increases not only the incidence and mortality rates of chronic disease, but also the medical and economic burden of disease [5-7]. Physical fitness has various degrees of influence on the activities and quality of life [1,8]. Poor physical fitness (below the 25th percentile of the fitness distribution) has a much greater impact on the risk of cardiovascular disease than insufficient physical activity [9]. Therefore, physical fitness needs to be considered as a fundamental assessment for people with a higher risk of chronic diseases.

Several physical fitness assessment methods have been established for reliability and validity. Physical fitness measures typically consist of cardiorespiratory fitness, muscle strength, endurance, agility, flexibility, and measures of body composition [1,10,11]. The 3-minute step test is one of the common cardiorespiratory fitness tests, consisting of stepping up and down a height of 23.0 cm-50.8 cm at a consistent step rate [12]. The 3-minute step test was shown to be reliable and valid in the general population and in patients with lung disease and rheumatoid arthritis [12-17]. Sufficient muscle power and endurance can reduce the risk of exercise injury and enhance cardiorespiratory capacity [18-20]. Wall squatting, push-up, and curl-up tests are common strength tests for the lower and upper extremities and abdominal muscles with established validity and reliability [11,21-23]. Balance and flexibility are important because poor stability may increase the risk of falls and limit functional activities [24-27]. Insufficient flexibility and mobility may restrict movement and cause pain [28-30]. Balance tests, the toe-touch test, the sit-and-reach test, and the Apley shoulder scratch test are common tests to assess balance, flexibility, and mobility [26,31-33]. However, most of these fitness tests are administered by a professional face to face, so patients or clients need to be present at a clinic or gym.

In consideration of cost and travel barriers, self-administered and home-based fitness tests may be more suitable for many people. However, there is currently little research on home-based fitness tests. One study showed that the home-based Senior Fitness Test, using inertia sensors and a depth camera, led to greater leg or arm strength, aerobic endurance, and flexibility [29]. The InterWalk Fitness Test incorporates indirect calorimetry and acceleration monitoring and was found to be accurate and reliable for persons with type 2 diabetes [34]. The self-administered Canadian Home Fitness Test was developed to assess cardiorespiratory endurance with a double 8-inch step and has an established record of safety and predictive ability [35-37]. Additional reliable home-based fitness tests that are easy to use and record data on accessible software platforms are needed.

As mobile technologies have advanced, an increasing number of health-related apps have emerged [38]. Some health apps provide patient education about lifestyle and health behaviors, some provide pain management, and others provide physical fitness assessments or interventions [34,38-40]. Among commercial fitness apps, most focus on cardiorespiratory fitness assessments, such as the submaximal walking data collected by a smartphone's accelerometer [40]. Some apps focus on functional performance, such as movement speed or leg strength during functional activities [41]. However, most commercial fitness apps lack supporting evidence [40]. Only a few fitness apps have been tested for validity and reliability, and most are rated as having moderate to good validity [34,42-44]. Insufficient monitoring of physiological signals (eg, heart rate) during cardiorespiratory fitness assessment was noted among the available apps [40]. Therefore, we designed R Plus Health (Recovery Plus Inc), a digital health app that incorporates all the usual components of a fitness assessment but also monitors physiological signals.

## Objective

The aim of this study is to investigate the test-retest reliability of home-based fitness assessments using the R Plus Health app in healthy young adults.

# Methods

## Participants

A total of 31 participants were recruited with convenience sampling from 4 departments of a technology company in Chengdu, China. Sampling was performed via random draw. The inclusion criteria were healthy adults with normal health examinations, aged between 18 and 75 years, and with the ability to use smartphones. Those who rated more than 3 out of 10 on the visual analog pain scale; had poor compliance or were not willing to cooperate with the assessment; had regular strengthening sessions during the study period; had a history of alcohol abuse or illegal drug use; were pregnant, lactating, or trying to become pregnant; had participated in other clinical trials within 3 months before this study; and had uncontrolled chronic diseases were excluded. The participants received oral and written information about the study, and informed consent was obtained from all participants. This study was approved by the Chinese Ethics Committee of Registering Clinical Trials (ChiCTR2000030905).

## R Plus Health App

The R Plus Health app was developed as a tool for healthy adults and patients with chronic diseases. It includes fitness assessments and individualized exercise prescriptions with physiological signal monitors (eg, heart rate monitor). After



XSL•FO
RenderX

downloading the R Plus Health app, the participants received an informed safety declaration and completed a health questionnaire, which was checked by doctors or other professional health care providers on the web. Through oral and video guidance, the participants were then instructed on how to perform the fitness assessments with maximal effort. The fitness assessments in the app included cardiorespiratory fitness, strength, balance, mobility, and flexibility tests (Figure 1). These fitness assessments have established clinical validity and reliability [12-25,30-33]. To complete the cardiorespiratory fitness test and record a real-time heart rate, the participants were required to wear a heart rate monitor below the sternum on a strap around the chest during testing (Figure 2). The heart rate monitor (Magene H64 dual protocol heart rate sensor) is compatible with the app and has Conformite Europeenne and Federal Communications Commission certification. Finally, according to the results of the fitness assessments and the overall health condition of each participant, a proper individualized exercise prescription was suggested by professional teams in the app.

**Figure 1.** Video demonstration of the push-up test.

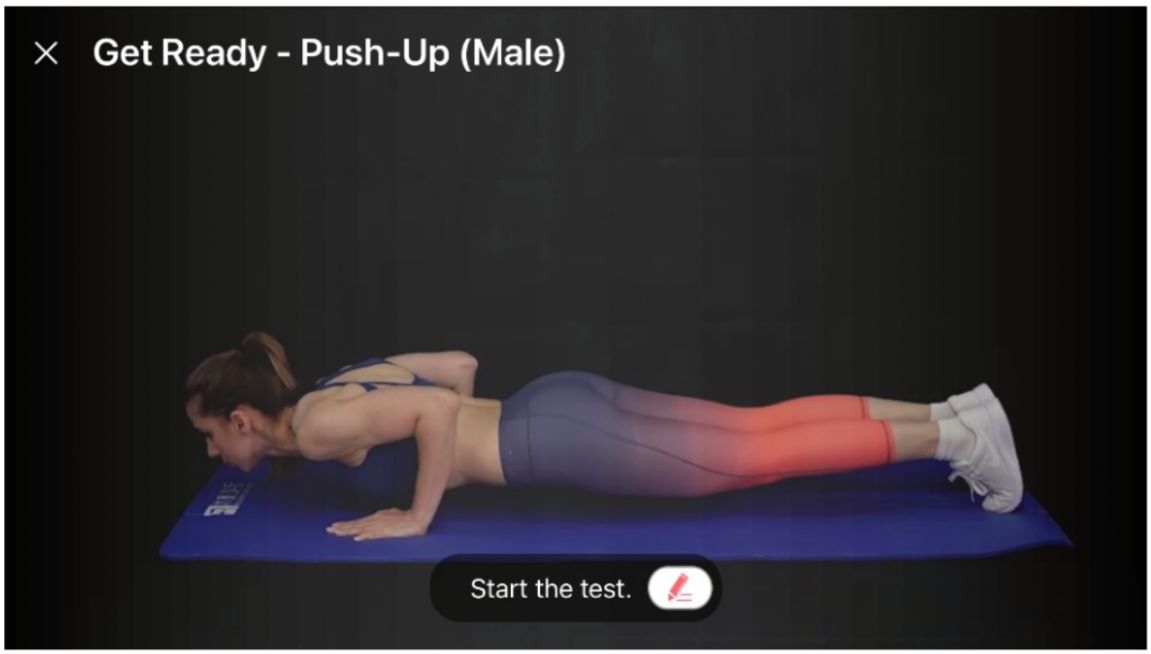



**Figure 2.** Demonstration of how to wear the heart rate monitor strap.

XSL•FO
RenderX

## Assessment Procedures

Eligible participants enrolled in the study, provided informed consent, downloaded the R Plus Health app, and filled in the health questionnaire with the assistance of a physiotherapist. The physiotherapist recorded the basic data, including pain level on a visual analog scale and the overall health condition of the participants, at the baseline and final assessments.

All participants self-administered 2 fitness assessments with a 2- to 3-day interval between assessments to provide the best reproducibility [45]. The fitness assessments were administered sequentially (cardiorespiratory endurance, strength, flexibility, mobility, and balance). The participants followed the guidance and instructions in the app for each fitness assessment.

The 3-minute step test measures cardiorespiratory endurance based on how quickly the heart rate returns to normal after a 3-minute step exercise [12,13]. First, the heart rate monitor strap was worn for a 5-minute rest period beside the 30-cm step (to establish a baseline). After watching the tutorial videos in the app, the participants stepped up and down at 96 beats per minute (bpm) using a metronome for a total of 3 minutes. After finishing the test, the participants rested for 1 minute. The participants could suspend the test if any discomfort occurred.

The push-up and curl-up tests measure muscle strength and endurance in the upper limbs and abdomen, respectively, based on the number of completed repetitions [11,21,23,46]. When performing the push-up test, there were 2 variations in the starting position. The standard push-up test involved having the knees off the ground in the push-up position and was used for male participants. The modified push-up test involved having the knees on the ground and was used for female participants. The participants performed as many push-ups as possible with the correct form within 40 seconds. The curl-up test began with the participants lying on their back, knees bent at approximately 90°, feet flat on the floor, and arms straight with the palms of their hands resting on their thighs. The participants curled up and down at 40 bpm using a metronome. If the participants could not continue or stopped for more than 5 seconds, they clicked the *completed* button and recorded the repetitions.

The wall squatting test measures muscle strength and endurance in the lower limbs based on the holding time [11,22]. The wall squatting test began with the participants in a standing position, feet shoulder width apart and back against the wall; then, both knees were bent at a 90° angle. The participants held this squatting position for as long as possible. When the participants were finished, they could click the *completed* button and record the total time. If the participants held the position for more than 150 seconds, the app finished the test automatically.

The sit-and-reach test measures the flexibility of the hamstrings and the lower back with a ruler based on the distance [11,30,32]. The participants sat on the floor with their legs straight and their heels in line with a ruler, hands stacked, and palms facing downward. They then reached forward as far as possible along the measuring line. After reaching forward, the participants recorded their distance in centimeters.

The Apley scratch test or the upper extremity (UE) multipattern test measures the mobility of the upper limbs based on the distance between the middle fingers [33]. There were 2 patterns of upper limb flexibility: shoulder flexion, abduction, and external rotation and shoulder extension, adduction, and internal rotation. The participants performed these 2 patterns of movement for each upper limb and recorded the distance between both middle fingers. The results were classified as above average, normal, or below average.

The one-leg stance test measures balance based on the holding time [26]. The participants stood on one leg, bent the other leg 15-20 cm off the ground with their eyes open and their arms beside the hips. The participants maintained their balance for as long as possible. If the participants lost balance, they clicked the *completed* button, and the time was recorded in the app automatically. If the participants maintained balance for more than 30 seconds, the app finished the test automatically.

To minimize possible diurnal variation in physical fitness, the participants were instructed to perform the 2 assessments at the same time of the day. They were asked to avoid resistance training and exhausting work between assessments to minimize the potential effects of fatigue. After each test, the participants immediately recorded the results on paper to avoid recall effects and then sent them to the researchers. The researchers concealed the data of the participants in an envelope for anonymity and encoded the names as numbers to protect the privacy of the participants.

## Outcome Measures

At the baseline assessment, the descriptive data, pain score, and health condition of the participants were evaluated by a physiotherapist. Descriptive data included age, sex, height, and weight. The pain level was assessed using a visual analog scale from 0 (no pain) to 10 (worst pain). Health condition was assessed using a health-related questionnaire in the app and by a physiotherapist.

The outcomes of each fitness assessment included the raw data and converted score. The raw data were recorded as bpm, repetitions, seconds, and an ordinal scale. The converted scores (0-100) were computed using the app through normative data and a self-established score conversion system on the basis of the raw data.

The participants recorded the heart rate in bpm as raw data, and the converted scores used the same units. If someone could not complete the 3-minute step test, the reason was noted [12]. In each cardiorespiratory fitness test, 2 measurements were made: the average resting heart rate during the 5-minute rest period and the 1-minute recovery heart rate after the 3-minute step test.

The outcomes of the push-up, wall squatting, and curl-up tests were recorded as completed repetitions and total time. The flexibility of the lower limbs and lower back was measured in centimeters from negative to positive values. The mobility of the upper limbs was classified as above average, normal, or below average. The one-leg stance test recorded the total time in seconds [26,31-33].

## Data Analysis

Statistical analyses were conducted using SPSS 20.0 software (IBM Inc). Descriptive statistics were presented in the form of



XSL•FO
RenderX

mean and SD, and the relative and absolute test-retest reliabilities of the fitness assessments were estimated separately.

### *Relative Reliability*

The relative reliability of the fitness assessments was calculated using the intraclass correlation coefficient (ICC) with a 2-way mixed model (type absolute agreement). On the basis of the 95% CIs of the ICC estimates, agreement was rated as poor (<0.5), moderate (between 0.5 and 0.75), good (between 0.75 and 0.9), or excellent (>0.9) [47].

### *Absolute Reliability*

The absolute reliability of the fitness assessments was evaluated using the SE of measurement (SEM), the smallest real difference (SRD), and Bland–Altman analyses [47,48]. The SEM expressed the measurement error variation between the assessments within a group and was calculated as $SD_{pooled}\times\sqrt{(1\text{-ICC})}$ [49]. In this formula, $SD_{pooled}$ indicates the pooled SD for the 2 assessments. The SRD is a measure of sensitivity to change, represented as the magnitude of the change detected at a certain CI [50]. The $SRD_{90}$ is defined as the SEM of the difference scores at a 90% confidence level and was calculated as $1.65\times\sqrt{2}\times SEM$ [48]. If the difference between the 2 assessments was greater than the SRD, it was interpreted as a real change. For all measurements, the smaller the SEM and $SRD_{90}$, the greater the reliability.

The Bland–Altman analyses and plots assessed the agreement or repeatability of the 2 assessments [49,51]. They estimated the mean and SD difference between the 2 assessments and established limits of agreement (LOA) within a 95% CI [51]. The 95% LOA was calculated as the mean difference$\pm(SD_{diff}\times1.96)$. The $SD_{diff}$ indicates the SD of the difference between 2 measurements [48]. The scatter plots show the relationship between the difference between the 2 assessments (y-axis) and the mean of the 2 assessments (x-axis). More point scattering within the 95% LOA, along with a smaller range between the 2 limits, indicated a higher agreement [52,53].

## Results

### Overview

The characteristics of the participants and the descriptive statistics of the fitness assessments at baseline are shown in Tables 1 and 2, respectively. The study enrolled 31 participants (Table 1), which exceeded the minimum sample size of 26 (effect size of 0.5 and power of 0.8) calculated using G*power 3.1 [54]. The average age was 27.25 (4.0) years, and they had negligible pain (mean 0.19 out of 10 on the visual analog scale), which did not worsen during testing.

**Table 1.** Characteristics of the participants (N=31).

| Characteristic | Values |
|---|---|
| Age (years), mean (SD) | 27.25 (4.0) |
| **Sex, n (%)** | |
| Female | 16 (52) |
| Male | 15 (48) |
| Height (cm), mean (SD) | 168.66 (7.61) |
| Weight (kg), mean (SD) | 60.23 (11.41) |
| BMI ($kg/m^2$), mean (SD) | 21.03 (2.75) |
| Health status | Normal health examination |
| Pain assessment (range 0-10), mean (SD) | 0.19 (0.65) |



XSL•FO
RenderX

**Table 2.** Fitness assessments of the participants at baseline (N=31).

| Domain and test items | Raw data, mean (SD) | Converted score[a], mean (SD) |
|---|---|---|
| **Cardiovascular fitness** | | |
| HR[b] at rest[c] (bpm[d]) | 74.81 (9.6) | 51.23 (16.5) |
| 1-minute HR after test[e] (bpm) | 92.26 (18.3) | 60.19 (19.1) |
| UE[f] strength: push-up (repetitions) | 12.94 (9.3) | 58.71 (18.3) |
| Abdominal strength: curl-up (repetitions) | 19.55 (13.7) | 51.29 (18.8) |
| LE[g] strength: wall squatting (seconds) | 63.03 (26.1) | 53.23 (18.1) |
| LE flexibility: sit-and-reach (centimeters) | 2.85 (14.2) | 57.74 (25.3) |
| **Balance ability** | | 93.23 (19.4) |
| Right leg stance (seconds) | 31.77 (14.4) | —[h] |
| Left leg stance (seconds) | 30.55 (9.9) | — |
| **UE mobility[i]** | | 65.48 (18.8) |
| **UE multipattern (above average), n (%)** | | — |
| Right UE | 23 (74) | |
| Left UE | 16 (52) | |
| **UE multipattern (normal), n (%)** | | — |
| Right UE | 3 (10) | |
| Left UE | 6 (19) | |
| **UE multipattern (below average), n (%)** | | — |
| Right UE | 5 (16) | |
| Left UE | 9 (29) | |

[a]Converted score (0-100) from raw data in the app using normative data.
[b]HR: heart rate.
[c]Resting heart rate measurement.
[d]bpm: beats per minute.
[e]Heart rate recovery 1 minute after the 3-minute step test.
[f]UE: upper extremity.
[g]LE: lower extremity.
[h]Not available; no converted score was calculated respectively because the scores were averaged in balance ability.
[i]The upper extremity multipattern test was categorized into 3 classes (above average, normal, and below average).

Table 2 shows the results of the baseline fitness assessments as raw data (mean [SD]) and converted score (0-100). At the baseline assessments, the average heart rate was 74.81 bpm at rest, and the 1-minute recovery heart rate was 92.26 bpm after the 3-minute step test. In the strength tests, the average number of completed repetitions was 12.94 push-ups and 19.55 curl-ups, and the average holding time for the squatting test was 63.03 seconds.

## Relative Reliability

Table 3 summarizes the test-retest reliability of all the fitness assessments. On the basis of the raw data, the ICCs for all tests were 0.8-0.99. On the basis of the converted scores, the ICCs for all tests were 0.69 to 0.99. In most tests, the 95% CI was >0.5.



XSL•FO
RenderX

**Table 3.** Test-retest reliability of the fitness assessments (N=31).

| Test items | ICC[a] for the raw data | ICC for the converted score |
|---|---|---|
| HR[b] at rest[c] | 0.80 (0.58-0.90) | 0.69 (0.34-0.85) |
| 1-minute HR after test[d] | 0.92 (0.84-0.96) | 0.82 (0.63-0.92) |
| UE strength[e] | 0.97 (0.94-0.99) | 0.97 (0.93-0.99) |
| Abdominal strength[f] | 0.98 (0.95-0.99) | 0.94 (0.87-0.97) |
| LE strength[g] | 0.93 (0.85-0.96) | 0.82 (0.63-0.92) |
| LE flexibility[h] | 0.89 (0.77-0.95) | 1 |
| UE mobility[i] | N/A[j] | 0.99 (0.98-0.99) |
| Right leg stance | 0.99 (0.98-0.99) | 0.75 (0.5-0.88) |
| Left leg stance | 0.89 (0.77-0.95) | 0.75 (0.5-0.88) |

[a]ICC: intraclass correlation coefficient (at a 95% CI).

[b]HR: heart rate.

[c]Resting heart rate measurement.

[d]1-minute HR after test: heart rate recovery 1 minute after the 3-minute step test.

[e]UE strength: upper extremity strength (push-up test).

[f]Curl-up test.

[g]LE strength: lower extremity strength (wall squatting test).

[h]LE flexibility: lower extremity flexibility (sit-and-reach test).

[i]UE mobility: upper extremity mobility (upper extremity multipattern test).

[j]N/A: not applicable; no intraclass correlation coefficient value was calculated because the raw data of the upper extremity mobility test was the percentage of participants, not a continuous variable.

## Absolute Reliability

The absolute reliability and Bland–Altman analyses are presented in Table 4. The SEM and $SRD_{90}$ were 1.44-6.91 and 3.36-16.11, respectively, across the different assessments. The mean differences in UE strength, lower extremity (LE) flexibility, and right leg balance tests were close to 0. The 95% CI of the mean difference contained 0, indicating no significant systematic error between the 2 assessments in strength (−6.28 to 3.89 in the LE strength test and −1.54 to 0.89 in the UE strength test), flexibility (−2.65 to 3.57 in the LE flexibility test), and balance tests (−1.75 to 0.07 in the right leg stance test and −5.58 to 0.93 in the left leg stance test).



XSL•FO
RenderX

**Table 4.** Absolute reliability of the fitness assessments in raw data.

| Raw data of test items | SEM[a] | $SRD_{90}$[b] | Bland–Altman analyses | | | | |
|---|---|---|---|---|---|---|---|
| | | | d[c] | $SD_{diff}$[d] | SE of d[e] | 95% CI of d | LOA[f] |
| HR at rest[g] (bpm[h]) | 4.28 | 9.99 | 5.61 | 5.58 | 1.00 | 3.57 to 7.66 | −5.32 to 16.55 |
| 1-minute HR after test[i] (bpm) | 5.18 | 12.08 | 7.19 | 6.91 | 1.24 | 4.66 to 9.73 | −6.34 to 20.73 |
| UE strength[j] (repetitions) | 1.61 | 3.76 | −0.32 | 3.31 | 0.59 | −1.54 to 0.89 | −6.81 to 6.17 |
| Abdominal strength[k] (repetitions) | 2.08 | 4.85 | −1.74 | 4.00 | 0.72 | −3.21 to −0.27 | −9.58 to 6.10 |
| LE strength[l] (s) | 6.91 | 16.11 | −1.19 | 13.87 | 2.49 | −6.28 to 3.89 | −28.38 to 25.99 |
| LE flexibility[m] (cm) | 4.71 | 10.99 | 0.46 | 8.47 | 1.52 | −2.65 to 3.57 | −16.14 to 17.06 |
| Right leg stance (s) | 1.44 | 3.36 | −0.84 | 2.48 | 0.45 | −1.75 to 0.07 | −5.70 to 4.02 |
| Left leg stance (s) | 3.28 | 7.66 | −2.32 | 8.87 | 1.59 | −5.58 to 0.93 | −19.70 to 15.06 |

[a]SEM: SE of measurement.

[b]$SRD_{90}$: smallest real difference at a 90% confidence level.

[c]d: mean difference between 2 trials.

[d]$SD_{diff}$: SD of mean difference.

[e]$SD_{diff}/\sqrt{n}$.

[f]LOA: limits of agreement ($d \pm [SD_{diff} \times 1.96]$).

[g]HR at rest: resting heart rate measurement.

[h]bpm: beats per minute.

[i]1-minute HR after test: heart rate recovery in 1 minute after the 3-minute step test.

[j]UE strength: upper extremity strength (push-up test).

[k]Abdominal strength: curl-up test.

[l]LE strength: lower extremity strength (wall squatting test).

[m]LE flexibility: lower extremity flexibility (sit-and-reach test).

Figures 3-10 show the Bland–Altman plots of the differences between the 2 measurements for all tests. Reference lines show mean differences between time 1 and time 2 (solid line) and 95% LOA for the mean difference (dotted lines). The differences for most of the tests were within the 95% CI. The LOA were −5.32 to 16.55 for the heart rate at rest and −6.34 to 20.73 for the 1-minute heart rate after test. The LOA were −6.81 to 6.17 in the UE strength test, −9.58 to 6.10 in the abdominal strength test, and −28.38 to 25.99 in the LE strength test. The LOA were −16.14 to 17.06 in the LE flexibility test, −5.70 to 4.02 in the right leg stance test, and −19.70 to 15.06 in the left leg stance test. There were at most 3 outliers in the 1-minute heart rate after, LE strength, and right leg stance tests.



XSL•FO
RenderX

**Figure 3.** The Bland–Altman plots of differences between the 2 measurements in heart rate at rest. HR: heart rate.

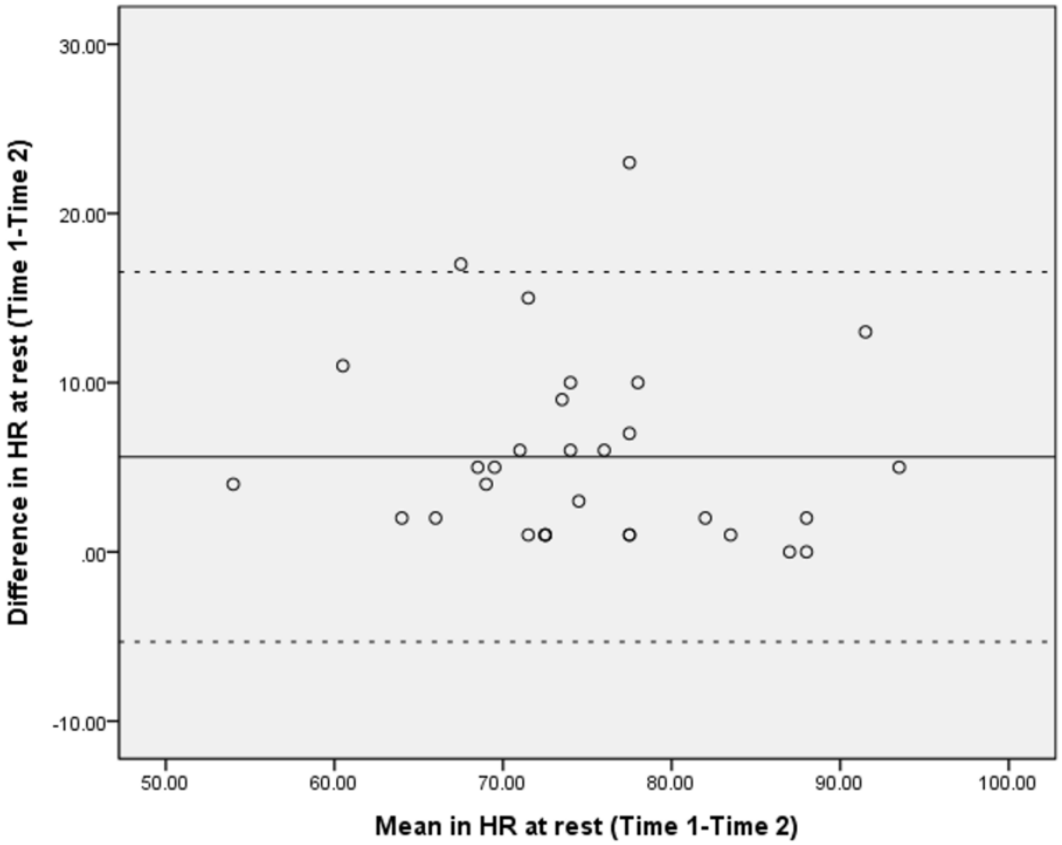
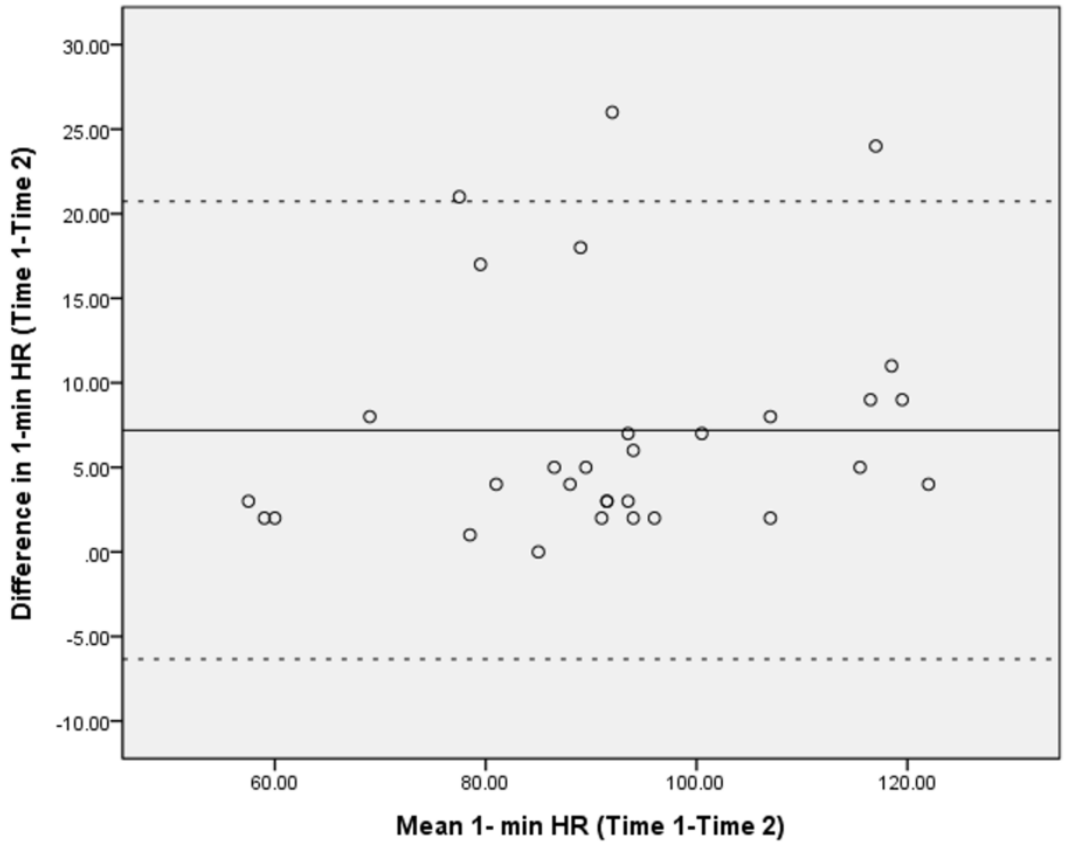



**Figure 4.** The Bland–Altman plots of differences between the 2 measurements in 1-minute heart rate recovery. HR: heart rate.






XSL•FO
RenderX

**Figure 5.** The Bland–Altman plots of differences between the 2 measurements in abdominal strength assessments. ab: abdominal.

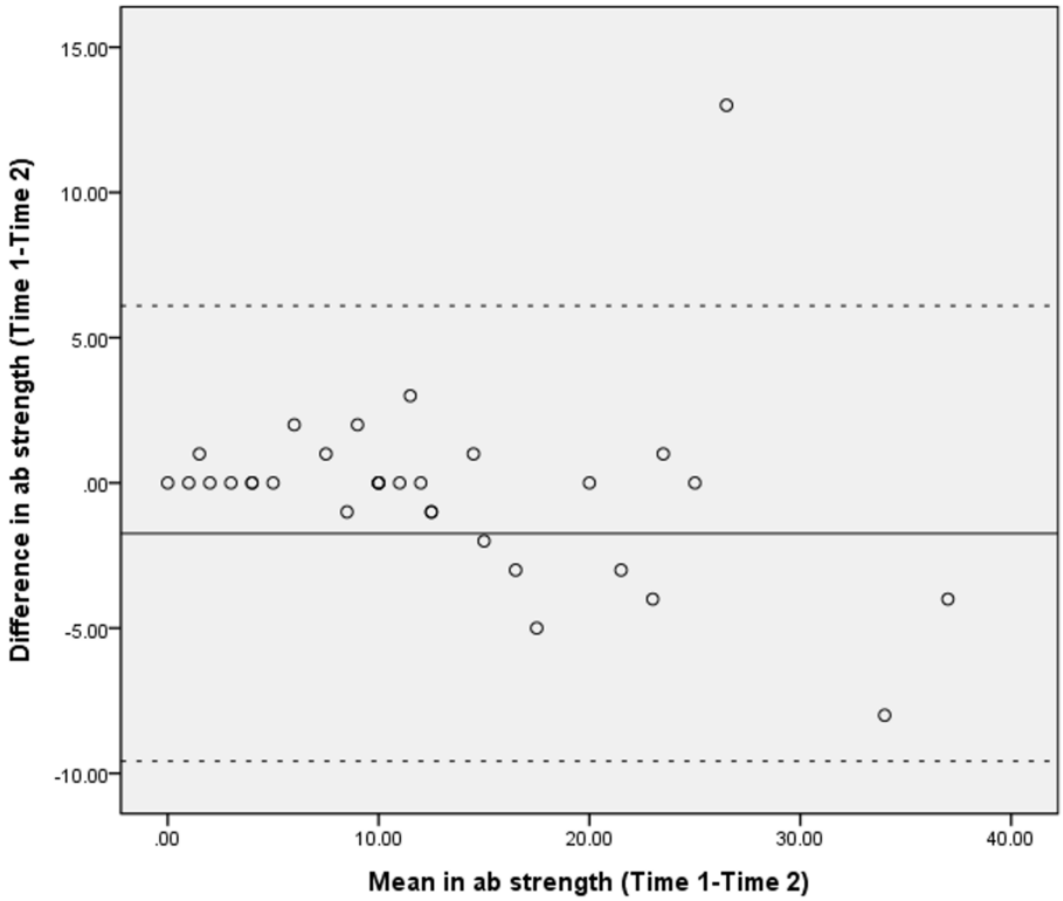
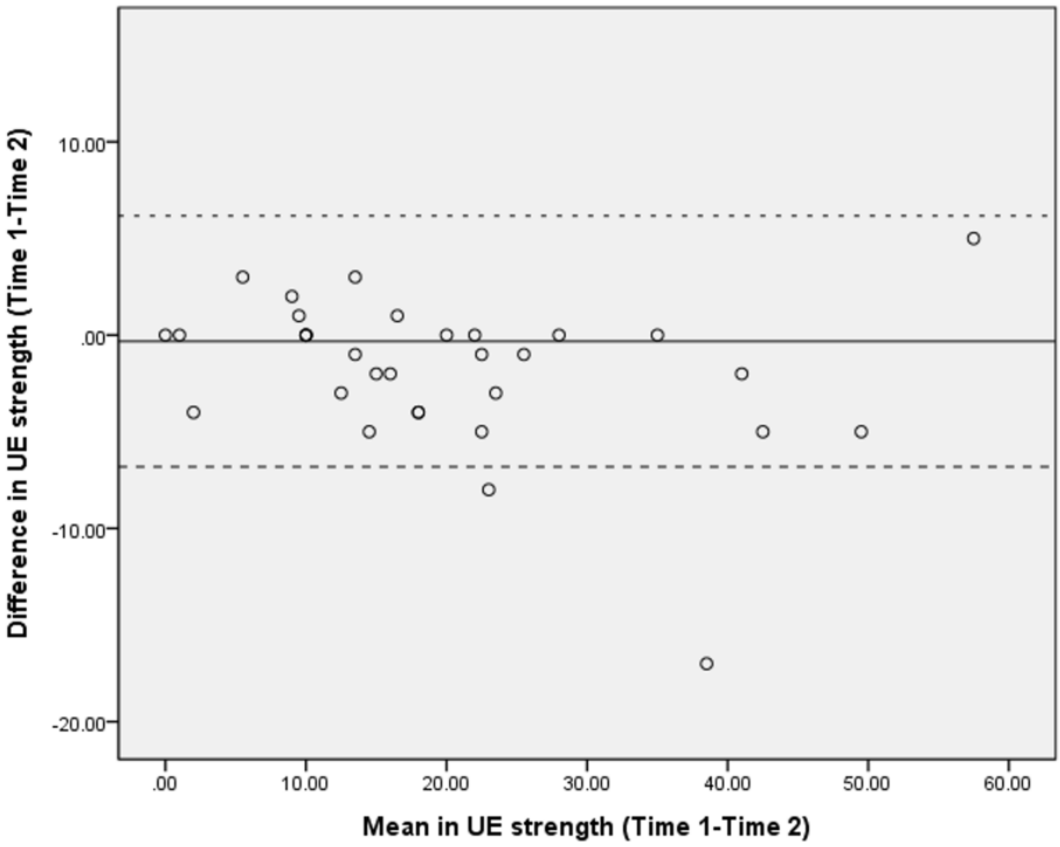



**Figure 6.** The Bland–Altman plots of differences between the 2 measurements in upper extremity strength assessments. UE: upper extremity.






XSL•FO
RenderX

**Figure 7.** The Bland–Altman plots of differences between the 2 measurements in lower extremity strength assessments. LE: lower extremity.

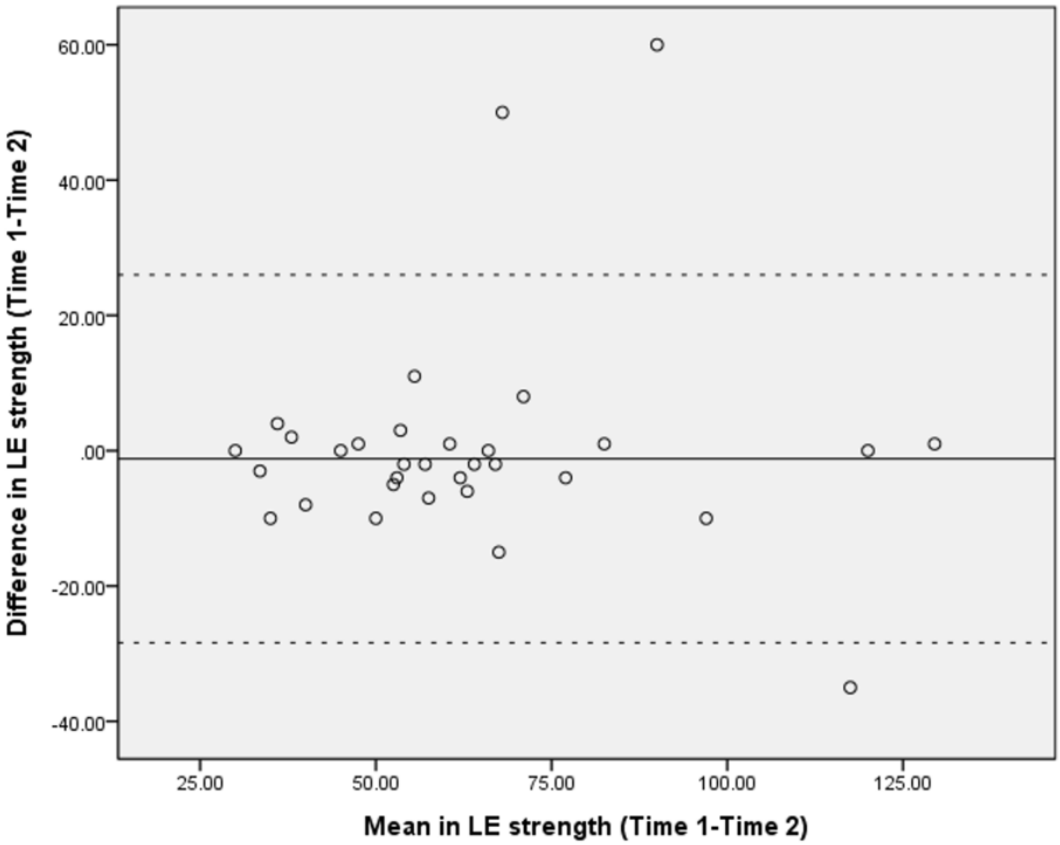
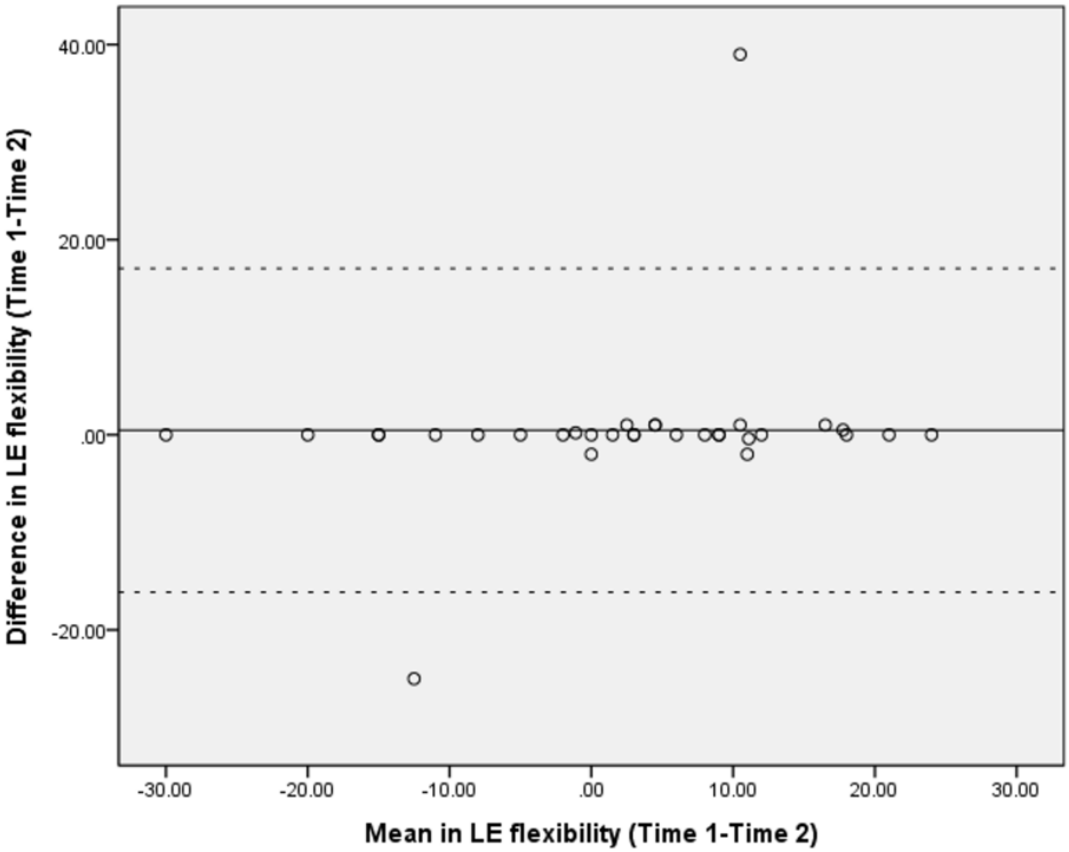



**Figure 8.** The Bland–Altman plots of differences between the 2 measurements in LE flexibility tests. LE: lower extremity.






XSL•FO
RenderX

**Figure 9.** The Bland–Altman plots of differences between the 2 measurements in right leg stance tests. R: right.

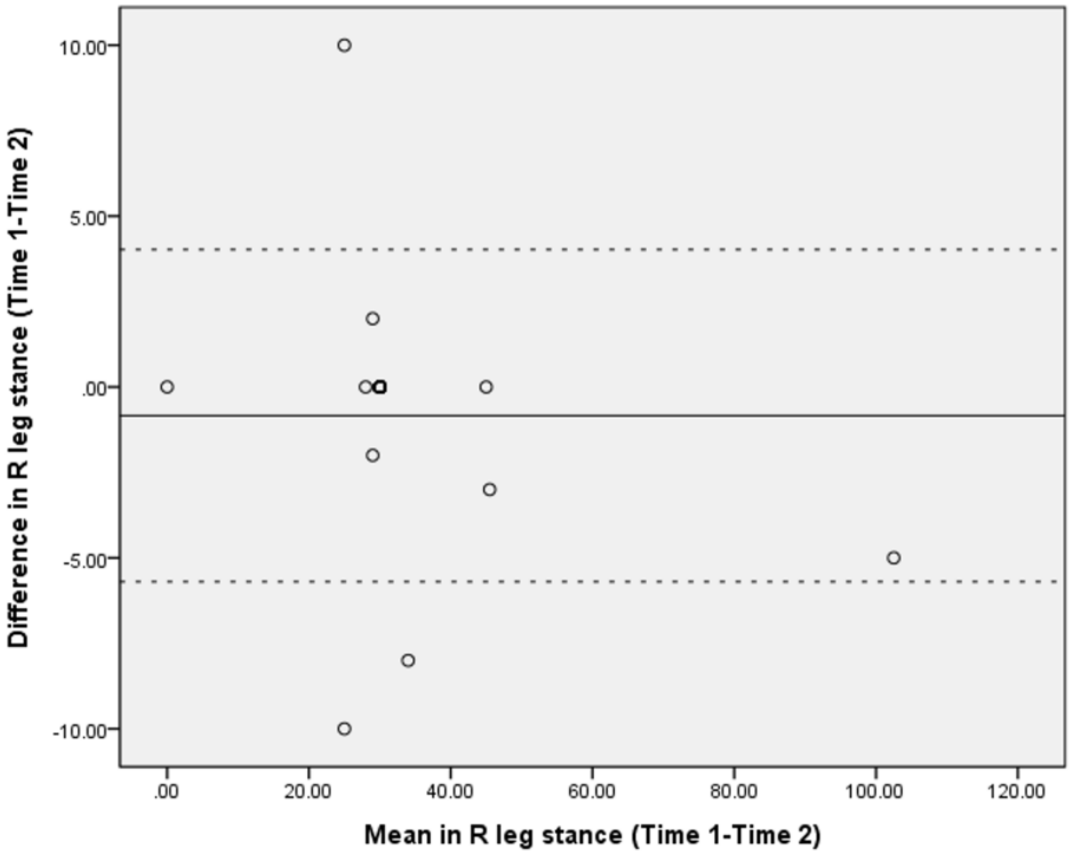
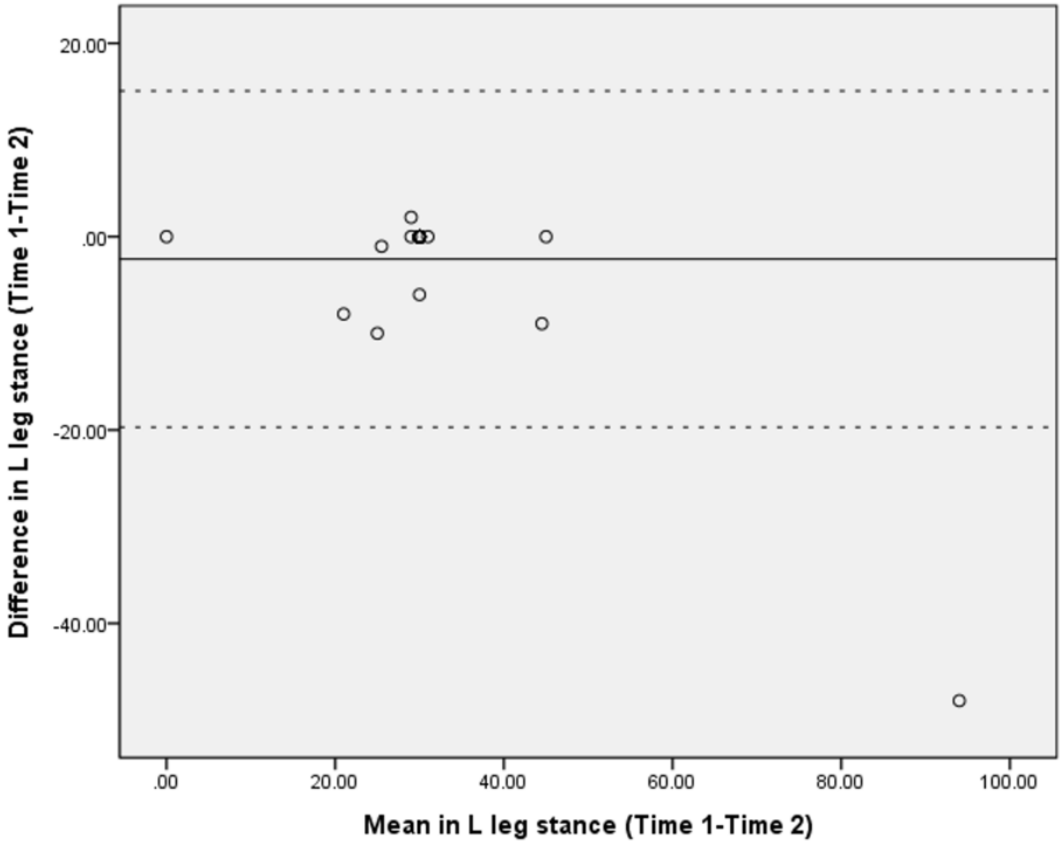



**Figure 10.** The Bland–Altman plots of differences between the 2 measurements in left leg stance tests. L: left.






XSL•FO
RenderX

## Discussion

### Principal Findings

This is the first study to investigate the test-retest reliability of home-based fitness assessments using a mobile health app in young, healthy adults. The results showed a moderate to good reliability of the fitness assessments. Therefore, through video and oral guidance, the app was shown to be reliable when applied to young users.

The self-administered fitness assessments in the app were feasible, with a low risk of injury. All participants completed the fitness assessments with the guidance of the R Plus Health app. Although some participants enrolled in the study had mild pain, they did not worsen after the fitness assessment. In other clinical research, it has been shown that mobile apps are able to conduct ecological momentary assessments, manage, and monitor patients with good adherence, detect symptoms, and evaluate the condition of a patient [55-57]. Therefore, well-designed mobile apps could offer a feasible means of self-assessment for clients and clinicians.

The results of our study were consistent with those of previous reliability studies [14]. Among the fitness assessments in the app, the test-retest reliabilities were moderate to good in this study. On the basis of the raw data, the ICCs for all tests were 0.8-0.99, indicating good to excellent reliability. On the basis of the converted scores, the ICCs for all tests were 0.68-0.99, indicating moderate to good reliability. The 95% CIs were above 0.5 in most tests. One previous study investigated the reliability of web-based versus supervised cardiovascular fitness assessments using the Young Men's Christian Association 3-minute step test for college students [14]. The results of that study showed that there were no significant differences in the recovery heart rate between the 2 groups and that self-assessed cardiovascular fitness measurements were reliable [14]. Another study investigating the reliability of the Chester Step Test in patients with chronic obstructive pulmonary disease showed good reliability (ICC>0.8) [58]. In a previous analysis of strength fitness tests, reliability was established in adolescents, with ICCs of 0.7-0.9 in push-up, curl-up, and wall squatting tests [59]. For balance tests, the ICCs of single-leg stance tests were found to be >0.77 in young adults using a computerized balance platform [60]. These findings suggest that, regardless of the methods of assessing fitness capacity (eg, web-based and supervised assessments), the use of standard procedures and precise guidance under signal monitoring can ensure an accurate measure of actual performance. Self-administered fitness assessments in the R Plus Health app can be one of these efficient and reliable methods.

In addition to the relative reliability, the absolute reliability can demonstrate the agreement and sensitivity of the mean differences between the assessments. The SRD is a measure of sensitivity to change and represents the magnitude of the change at a certain confidence level [50]. If the difference between 2 assessments was larger than the SRD, it could be considered a real change, and the smaller the SEM and $SRD_{90}$ of the difference, the greater the reliability. In this study, the SEM and $SRD_{90}$ ranged from 1.44-6.91 and 3.36-16.11, respectively. For example, if the change was more than 16.11 in the wall squatting test, it was considered real at a 90% confidence level. In this study, the $SRD_{90}$ in the wall squatting, push-up, and curl-up tests were 16.11, 3.76, and 4.85, respectively. These values were greater than the between-assessment changes reported in a previous study, which were 6.2, 2.6, and 0.1 for the wall squatting, push-up, and curl-up tests, respectively [59]. Different results might be because of different populations, ages, and experimental designs.

The Bland–Altman analyses and plots were generated to measure the repeatability of 2 measurement systems or of several trials using one method [49,51]. The scattering of data points within the 95% LOA and a smaller range between the 2 limits indicated higher agreement [52,53]. The 95% CI of the mean difference contained 0, indicating no significant systematic error between the 2 assessments for the strength, flexibility, and balance tests. The range of the LOA was slightly narrower in the UE strength (−6.81 to 6.17), abdominal muscle strength (−9.58 to 6.10), and right leg stance (−5.70 to 4.02) tests, indicating higher agreement. There was at least one outlier in each fitness assessment, and at most 3 outliers (in the 1-minute heart rate, LE strength, and right leg stance tests), which might be due to familiarization or fatigue in the second test.

Sufficient physical fitness is critical in daily life. It can decrease the risk of cardiovascular disease, pain, and injuries and improve the performance of life activities [9,18-20]. From the results of the cardiorespiratory fitness assessments in this study, the mean heart rate after 1-minute recovery from the step test was 92.26 bpm, indicating above average fitness base on the normative data [61]. In the LE strength wall squatting test, the mean holding time was 63.03 seconds, indicating an average fitness level [62]. In the LE flexibility sit-and-reach test, the mean distance was 2.85 cm, categorized as an average fitness level [11]. Even though the enrolled participants were generally in good health, they were below average in some of the fitness components. In the push-up test for UE strength and the curl-up test for abdominal muscle strength test, the mean number of repetitions was 12.94 and 19.55, respectively, which was below average based on the normative data and showed a need for improvement [63,64]. In the single-leg stance balance test, the mean holding time was 30 seconds (31.77 seconds for the right leg and 30.55 seconds for the left leg), indicating a below average fitness level [65]. Lack of muscle strength and balance can increase the risk of falls, pain, and injuries and limit daily life activities [24-27]. Therefore, comprehensive fitness assessments are essential.

Each participant differed in their performance in the physical fitness assessments according to variable self-conditions between the 2 assessments, and the results also differed from one participant to another. Even in healthy participants without chronic diseases, mild pain can lead to low strength in the extremities. Pain can inhibit muscle firing, and the lack of muscle contraction can decrease the stability of the joints and in turn, produce pain [66]. In other situations, insufficient muscle strength can lead to poor cardiorespiratory fitness. Evidence has shown that muscular fitness is related to cardiovascular prognosis and mortality [67]. As a result, according to the



XSL•FO
RenderX

individual situation, it is important to detect weaknesses in the fitness profile and provide proper assessments and advice to clients. Through a comprehensive fitness assessment composed of cardiovascular endurance, strength, flexibility, and balance tests, the R Plus Health app can provide clinicians with a complete picture of the clients' fitness. Clinicians can then choose to provide other detailed assessments on the web or at the clinic, which not only increases the efficiency of the evaluations but also decreases the medical and economic burden.

### Limitations and Future Studies

This study had several limitations. First, the level of difficulty in similar assessments differed from one participant to another, which may lead to ceiling or floor effects. According to the individual situation, automatic adjustment of the grade of assessments will be essential. Second, it is difficult to ascertain the accuracy of the performance assessments in the app without professional supervision. That is, the results of the app might not be identical to testing under professional supervision. In this study design, the results from different testing situations (with or without supervision) could not be compared. One solution to this problem would be to apply suitable monitoring equipment (eg, motion capture analysis and artificial intelligence techniques) to increase the precision of the assessments. However, this creates an additional technological burden. Cross-validation of the outcomes collected by the app versus professional staff will be the subject of future studies. Third, the study recruited young, healthy adults, so the results of the fitness assessments should not be generalized to other populations, such as older adults or patients with chronic diseases. Therefore, the fitness assessments in the app need to be conducted in other populations to compare the results between the app and clinical testing. Testing of the R Plus Health app in additional populations will be conducted in the future.

### Conclusions

Home-based fitness assessments using a mobile health app were reliable and feasible in young, healthy adults. The results showed moderate to good reliability, and the testing process caused negligible pain effects. This study highlighted an important contribution of mobile health apps to health care, that is, that healthy adults can self-administer fitness tests and thereby reduce overall costs. The results of mobile fitness assessments can offer a reliable understanding of a person's health condition and help prescribe a safe and suitable exercise training regimen. Expansion of the use of this technology to different populations (eg, patients with chronic diseases or users with poor fitness) will offer widespread benefits to both patients and the health care system.

### Acknowledgments

This research was partially supported by the Ministry of Science and Technology (MOST-109-2314-B-182-030, and 110-2314-B-182-017) and Chang Gung Memorial Hospital (CMRPD1I0141 and CMRPD 1I0142) in Taiwan.

The authors thank Si-jing Ye and Jing Wei for data collection and Chong Jiang for assistance in conducting the survey. All authors read and approved the final manuscript.

### Conflicts of Interest

None declared.

### References

1. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131 [FREE Full text] [Medline: 3920711]
2. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. Can Med Asso J 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
3. Katzmarzyk PT. Physical activity, sedentary behavior, and health: paradigm paralysis or paradigm shift? Diabetes 2010 Nov 27;59(11):2717-2725 [FREE Full text] [doi: 10.2337/db10-0822] [Medline: 20980470]
4. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):1077-1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
5. Chronic diseases and their common risk factors. World Health Organization. 2005. URL: https://www.who.int/chp/chronic_disease_report/media/Factsheet1.pdf [accessed 2021-11-02]
6. Anderson E, Durstine J. Physical activity, exercise, and chronic diseases: a brief review. Sport Med Health Sci 2019 Dec;1(1):3-10 [FREE Full text] [doi: 10.1016/j.smhs.2019.08.006]
7. Yin P, Qi J, Liu Y, Liu J, Li J, Zeng X, et al. Burden of disease in the Chinese population from 2005 to 2017. Chin Circ J 2019;34(12):1146-1147 [FREE Full text]
8. Weening-Dijksterhuis E, de Greef MH, Scherder EJ, Slaets JP, van der Schans CP. Frail institutionalized older persons: a comprehensive review on physical exercise, physical fitness, activities of daily living, and quality-of-life. Am J Phys Med Rehabil 2011 Feb;90(2):156-168. [doi: 10.1097/PHM.0b013e3181f703ef] [Medline: 20881587]
9. Williams PT. Physical fitness and activity as separate heart disease risk factors: a meta-analysis. Med Sci Sports Exerc 2001 May;33(5):754-761 [FREE Full text] [doi: 10.1097/00005768-200105000-00012] [Medline: 11323544]



XSL•FO
RenderX

10. Wilder RP, Greene JA, Winters KL, Long WB, Gubler K, Edlich RF. Physical fitness assessment: an update. J Long Term Eff Med Implants 2006;16(2):193-204. [doi: 10.1615/jlongtermeffmedimplants.v16.i2.90] [Medline: 16700660]
11. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. Philadelphia, PA: Lippincott Williams & Wilkins; 2013.
12. Andrade C, Cianci R, Malaguti C, Corso SD. The use of step tests for the assessment of exercise capacity in healthy subjects and in patients with chronic lung disease. J Bras Pneumol 2012;38(1):116-124 [FREE Full text] [doi: 10.1590/s1806-37132012000100016] [Medline: 22407048]
13. Lee O, Kim SS, Kim YS, Son HJ, Kim YM, Choi BY. Correlation between YMCA step-test and maximum oxygen consumption (VO2max) as measurement tools for cardiorespiratory. Korean J Epidemiol 2008 Jun 30;30(1):73-81. [doi: 10.4178/kje.2008.30.1.73]
14. Liguori G, Mozumdar A. Reliability of self assessments for a cardiovascular fitness assessment. Int J Fitness 2009;5(1):33-40 [FREE Full text]
15. Cooney JK, Moore JP, Ahmad YA, Jones JG, Lemmey AB, Casanova F, et al. A simple step test to estimate cardio-respiratory fitness levels of rheumatoid arthritis patients in a clinical setting. Int J Rheumatol 2013;2013:174541-174548 [FREE Full text] [doi: 10.1155/2013/174541] [Medline: 24454385]
16. Bennett H, Parfitt G, Davison K, Eston R. Validity of submaximal step tests to estimate maximal oxygen uptake in healthy adults. Sports Med 2016 May 15;46(5):737-750. [doi: 10.1007/s40279-015-0445-1] [Medline: 26670455]
17. Teren A, Zachariae S, Beutner F, Ubrich R, Sandri M, Engel C, et al. Incremental value of Veterans Specific Activity Questionnaire and the YMCA-step test for the assessment of cardiorespiratory fitness in population-based studies. Eur J Prev Cardiol 2016 Jul;23(11):1221-1227. [doi: 10.1177/2047487315621844] [Medline: 26656282]
18. Oliver GD, Adams-Blair HR. Improving core strength to prevent injury. J Phys Edu Recreat Dance 2010 Sep;81(7):15-19. [doi: 10.1080/07303084.2010.10598503]
19. Cho KH, Bok SK, Kim Y, Hwang SL. Effect of lower limb strength on falls and balance of the elderly. Ann Rehabil Med 2012 Jun;36(3):386-393 [FREE Full text] [doi: 10.5535/arm.2012.36.3.386] [Medline: 22837975]
20. Watanabe M, Matsumoto T, Ono S, Koseki H, Watarai K. Relationship of lower extremity alignment during the wall squat and single-leg jump: assessment of single-leg landing using three-dimensional motion analysis. J Phys Ther Sci 2016 Jun;28(6):1676-1680 [FREE Full text] [doi: 10.1589/jpts.28.1676] [Medline: 27390393]
21. Hall GL, Hetzler RK, Perrin D, Weltman A. Relationship of timed sit-up tests to isokinetic abdominal strength. Res Q Exerc Sport 1992 Mar;63(1):80-84. [doi: 10.1080/02701367.1992.10607560] [Medline: 1574665]
22. Blazevich AJ, Gill N, Newton RU. Reliability and validity of two isometric squat tests. J Strength Cond Res 2002 May;16(2):298-304. [Medline: 11991785]
23. Cogley RM, Archambault TA, Fibeger JF, Koverman MM, Youdas JW, Hollman JH. Comparison of muscle activation using various hand positions during the push-up exercise. J Strength Condit Res 2005;19(3):628-633. [doi: 10.1519/00124278-200508000-00024]
24. Harris J, Eng J, Marigold D, Tokuno C, Louis C. Relationship of balance and mobility to fall incidence in people with chronic stroke. Phys Ther 2005;85(2):150-158. [doi: 10.1093/ptj/85.2.150]
25. Hrysomallis C. Relationship between balance ability, training and sports injury risk. Sports Med 2007;37(6):547-556. [doi: 10.2165/00007256-200737060-00007] [Medline: 17503879]
26. Schepens S, Goldberg A, Wallace M. The short version of the Activities-specific Balance Confidence (ABC) scale: its validity, reliability, and relationship to balance impairment and falls in older adults. Arch Gerontol Geriatr 2010;51(1):9-12 [FREE Full text] [doi: 10.1016/j.archger.2009.06.003] [Medline: 19615762]
27. Urrunaga-Pastor D, Moncada-Mapelli E, Runzer-Colmenares F, Bailon-Valdez Z, Samper-Ternent R, Rodriguez-Mañas L, et al. Factors associated with poor balance ability in older adults of nine high-altitude communities. Arch Gerontol Geriatr 2018;77:108-114 [FREE Full text] [doi: 10.1016/j.archger.2018.04.013] [Medline: 29738899]
28. Malliaras P, Hogan A, Nawrocki A, Crossley K, Schache A. Hip flexibility and strength measures: reliability and association with athletic groin pain. Br J Sports Med 2009 Oct 11;43(10):739-744. [doi: 10.1136/bjsm.2008.055749] [Medline: 19282303]
29. Kelley MJ, Shaffer MA, Kuhn JE, Michener LA, Seitz AL, Uhl TL, et al. Shoulder pain and mobility deficits: adhesive capsulitis. J Orthop Sports Phys Ther 2013 May;43(5):1-31. [doi: 10.2519/jospt.2013.43.1.A1] [Medline: 23636125]
30. Mistry G, Vyas N, Sheth M. Comparison of hamstrings flexibility in subjects with chronic low back pain versus normal individuals. J Clin Exp Res 2014;2(1):85-88. [doi: 10.5455/jcer.201413]
31. Ayala F, de Baranda PS, De Ste Croix M, Santonja F. Reproducibility and criterion-related validity of the sit and reach test and toe touch test for estimating hamstring flexibility in recreationally active young adults. Phys Ther Sport 2012 Nov;13(4):219-226 [FREE Full text] [doi: 10.1016/j.ptsp.2011.11.001] [Medline: 23068896]
32. Mayorga-Vega D, Viciana J, Cocca A, Merino-Marban R. Criterion-related validity of toe-touch test for estimating hamstring extensibility: a meta-analysis. J Hum Sport Exerc 2014 Jul;9(1):188-200. [doi: 10.4100/jhse.2014.91.18]
33. Sprague PA, Mokha GM, Gatens DR, Rodriguez R. The relationship between glenohumeral joint total rotational range of motion and the functional movement screen™ shoulder mobility test. Int J Sports Phys Ther 2014 Oct;9(5):657-664 [FREE Full text] [Medline: 25328828]




XSL•FO
RenderX

34. Brinkløv CF, Thorsen IK, Karstoft K, Brøns C, Valentiner L, Langberg H, et al. Criterion validity and reliability of a smartphone delivered sub-maximal fitness test for people with type 2 diabetes. BMC Sports Sci Med Rehabil 2016;8:31 [FREE Full text] [doi: 10.1186/s13102-016-0056-7] [Medline: 28174664]
35. Jetté M, Campbell J, Mongeon J, Routhier R. The Canadian Home Fitness Test as a predictor for aerobic capacity. Can Med Assoc J 1976 Apr 17;114(8):680-682 [FREE Full text] [Medline: 1260614]
36. Shephard RJ, Bailey DA, Mirwald RL. Development of the Canadian Home Fitness Test. Can Med Assoc J 1976 Apr 17;114(8):675-679 [FREE Full text] [Medline: 56979]
37. Shephard RJ. The current status of the Canadian Home Fitness Test. Br J Sports Med 1980 Jul 01;14(2-3):114-125 [FREE Full text] [doi: 10.1136/bjsm.14.2-3.114] [Medline: 7407449]
38. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
39. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
40. Muntaner-Mas A, Martinez-Nicolas A, Lavie CJ, Blair SN, Ross R, Arena R, et al. A systematic review of fitness apps and their potential clinical and sports utility for objective and remote assessment of cardiorespiratory fitness. Sports Med 2019 Apr;49(4):587-600 [FREE Full text] [doi: 10.1007/s40279-019-01084-y] [Medline: 30825094]
41. Ruiz-Cárdenas JD, Rodríguez-Juan JJ, Smart R, Jakobi J, Jones G. Validity and reliability of an iPhone App to assess time, velocity and leg power during a sit-to-stand functional performance test. Gait Posture 2018 Jan;59:261-266 [FREE Full text] [doi: 10.1016/j.gaitpost.2017.10.029] [Medline: 29102856]
42. Capela NA, Lemaire ED, Baddour NC. A smartphone approach for the 2 and 6-minute walk test. Annu Int Conf IEEE Eng Med Biol Soc 2014;2014:958-961. [doi: 10.1109/EMBC.2014.6943751] [Medline: 25570119]
43. Brooks GC, Vittinghoff E, Iyer S, Tandon D, Kuhar P, Madsen KA, et al. Accuracy and usability of a self-administered 6-minute walk test smartphone application. Circ Heart Fail 2015 Sep;8(5):905-913 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.115.002062] [Medline: 26283292]
44. Altini M, Van Hoof C, Amft O. Relation between estimated cardiorespiratory fitness and running performance in free-living: an analysis of HRV4Training data. In: Proceedings of the IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). 2017 Presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); Feb. 16-19, 2017; Orlando, FL, USA p. 16-19. [doi: 10.1109/bhi.2017.7897252]
45. Monteiro ER, Vingren JL, Neto VG, Neves EB, Steele J, Novaes JS. Effects of different between test rest intervals in reproducibility of the 10-repetition maximum load test: a pilot study with recreationally resistance trained men. Int J Exerc Sci 2019;12(4):932-940 [FREE Full text] [Medline: 31523350]
46. Ferguson B. ACSM's Guidelines for Exercise Testing and Prescription 9th Ed. 2014. J Can Chiropr Assoc 2014;58(3):328 [FREE Full text]
47. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
48. Chuang L, Wu C, Lin K, Hsieh C. Relative and absolute reliability of a vertical numerical pain rating scale supplemented with a faces pain scale after stroke. Phys Ther 2014 Jan;94(1):129-138. [doi: 10.2522/ptj.20120422] [Medline: 24029301]
49. Bruton A, Conway J, Holgate S. Reliability: what is it, and how is it measured? Physiotherapy 2000 Feb;86(2):94-99 [FREE Full text] [doi: 10.1016/s0031-9406(05)61211-4]
50. Schuck P, Zwingmann C. The 'smallest real difference' as a measure of sensitivity to change: a critical analysis. Int J Rehabil Res 2003 Jun;26(2):85-91. [doi: 10.1097/00004356-200306000-00002] [Medline: 12799601]
51. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25(2):141-151 [FREE Full text] [doi: 10.11613/BM.2015.015] [Medline: 26110027]
52. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 08;1(8476):307-310. [Medline: 2868172]
53. Myles P, Cui J. Using the Bland-Altman method to measure agreement with repeated measures. Br J Anaesth 2007 Sep;99(3):309-311 [FREE Full text] [doi: 10.1093/bja/aem214] [Medline: 17702826]
54. Buchner A, Erdfelder E, Faul F, Lang AG. G*Power : statistical power analyses for Mac and Windows. Heinrich-Heine-Universität Düsseldorf. 2001. URL: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower [accessed 2021-09-08]
55. Holtz B, Whitten P. Managing asthma with mobile phones: a feasibility study. Telemed J E Health 2009 Nov;15(9):907-909. [doi: 10.1089/tmj.2009.0048] [Medline: 19919198]
56. Juengst SB, Graham KM, Pulantara IW, McCue M, Whyte EM, Dicianno BE, et al. Pilot feasibility of an mHealth system for conducting ecological momentary assessment of mood-related symptoms following traumatic brain injury. Brain Inj 2015 Aug;29(11):1351-1361. [doi: 10.3109/02699052.2015.1045031] [Medline: 26287756]
57. Lee J, Song S, Ahn JS, Kim Y, Lee JE. Use of a mobile application for self-monitoring dietary intake: feasibility test and an intervention study. Nutrients 2017 Jul 13;9(7):748 [FREE Full text] [doi: 10.3390/nu9070748] [Medline: 28703758]




XSL•FO
RenderX

58. de Camargo AA, Justino T, de Andrade CH, Malaguti C, Corso SD. Chester step test in patients with COPD: reliability and correlation with pulmonary function test results. Respir Care 2011 Jul 01;56(7):995-1001 [FREE Full text] [doi: 10.4187/respcare.01047] [Medline: 21740727]
59. Lubans DR, Morgan P, Callister R, Plotnikoff RC, Eather N, Riley N, et al. Test-retest reliability of a battery of field-based health-related fitness measures for adolescents. J Sports Sci 2011 Apr;29(7):685-693. [doi: 10.1080/02640414.2010.551215] [Medline: 21391082]
60. Muehlbauer T, Roth R, Mueller S, Granacher U. Intra and intersession reliability of balance measures during one-leg standing in young adults. J Strength Cond Res 2011 Aug;25(8):2228-2234. [doi: 10.1519/JSC.0b013e3181fb393b] [Medline: 21685809]
61. YMCA of the USA. YMCA Fitness Testing and Assessment Manual. Champaign, Illinois: Human Kinetics Publishers; 2000:1-250.
62. McIntosh G, Wilson L. Trunk and lower extremity muscle endurance: normative data for adults. J Rehabil Outcomes Meas 1998;2(4):20-39 [FREE Full text]
63. Faulkner R, Sprigings E, McQuarrie A, Bell RD. A partial curl-up protocol for adults based on an analysis of two procedures. Can J Sport Sci 1989 Sep;14(3):135-141. [Medline: 2531031]
64. Canadian Society for Exercise Physiology. Canadian Physical Activity, Fitness & Lifestyle Approach : CSEP - Health & Fitness Program's Health-related Appraisal & Counselling Strategy. Ottawa, Ont: Canadian Society for Exercise Physiology; 2004.
65. Springer BA, Marin R, Cyhan T, Roberts H, Gill NW. Normative values for the unipedal stance test with eyes open and closed. J Geriatr Phys Ther 2007;30(1):8-15. [doi: 10.1519/00139143-200704000-00003] [Medline: 19839175]
66. Verbunt JA, Seelen HA, Vlaeyen JW, Bousema EJ, van der Heijden GJ, Heuts PH, et al. Pain-related factors contributing to muscle inhibition in patients with chronic low back pain: an experimental investigation based on superimposed electrical stimulation. Clin J Pain 2005;21(3):232-240. [doi: 10.1097/00002508-200505000-00006] [Medline: 15818075]
67. Artero EG, Lee D, Lavie CJ, España-Romero V, Sui X, Church TS, et al. Effects of muscular strength on cardiovascular risk factors and prognosis. J Cardiopulm Rehabil Prev 2012;32(6):351-358 [FREE Full text] [doi: 10.1097/HCR.0b013e3182642688] [Medline: 22885613]

## Abbreviations

**bpm:** beats per minute
**ICC:** intraclass correlation coefficient
**LE:** lower extremity
**LOA:** limits of agreement
**SEM:** SE of measurement
**UE:** upper extremity

*Edited by G Eysenbach; submitted 18.02.21; peer-reviewed by C Jacob, G Fico, C Reis; comments to author 04.08.21; revised version received 16.09.21; accepted 12.10.21; published 08.12.21.*

*Please cite as:*
*Lin II, Chen YL, Chuang LL*
*Test-Retest Reliability of Home-Based Fitness Assessments Using a Mobile App (R Plus Health) in Healthy Adults: Prospective Quantitative Study*
*JMIR Form Res 2021;5(12):e28040*
*URL: https://formative.jmir.org/2021/12/e28040*
*doi:10.2196/28040*
*PMID:34657835*

©I-I Lin, You-Lin Chen, Li-Ling Chuang. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Adoption and Appropriateness of mHealth for Weight Management in the Real World: A Qualitative Investigation of Patient Perspectives

Jessica Y Breland[1], PhD; Khizran Agha[1], MBBS; Rakshitha Mohankumar[1], BA

Center for Innovation to Implementation, VA Palo Alto Health Care System, Menlo Park, CA, United States

**Corresponding Author:**
Jessica Y Breland, PhD
Center for Innovation to Implementation
VA Palo Alto Health Care System
795 Willow Road (MPD-152)
Menlo Park, CA, 94025
United States
Phone: 1 650 493 5000
Email: jessica.breland@va.gov

## Abstract

**Background:** Mobile health (mHealth) interventions for weight management can result in weight loss outcomes comparable to in-person treatments. However, there is little information on implementing these treatments in real-world settings.

**Objective:** This work aimed to answer two implementation research questions related to mHealth for weight management: (1) what are barriers and facilitators to mHealth adoption (initial use) and engagement (continued use)? and (2) what are patient beliefs about the appropriateness (ie, perceived fit, relevance, or compatibility) of mHealth for weight management?

**Methods:** We conducted semistructured interviews with patients with obesity at a single facility in an integrated health care system (the Veterans Health Administration). All participants had been referred to a new mHealth program, which included access to a live coach. We performed a rapid qualitative analysis of interviews to identify themes related to the adoption of, engagement with, and appropriateness of mHealth for weight management.

**Results:** We interviewed 24 veterans, seven of whom used the mHealth program. Almost all participants were ≥45 years of age and two-thirds were White. Rapid analysis identified three themes: (1) coaching both facilitates and prevents mHealth adoption and engagement by promoting accountability but leading to guilt among those not meeting goals; (2) preferences regarding the mode of treatment delivery, usability, and treatment content were barriers to mHealth appropriateness and adoption, including preferences for in-person care and a dislike of self-monitoring; and (3) a single invitation was not sufficient to facilitate adoption of a new mHealth program. Themes were unrelated to participants' age, race, or ethnicity.

**Conclusions:** In a study assessing real-world use of mHealth in a group of middle-aged and older adults, we found that—despite free access to mHealth with a live coach—most did not complete the registration process. Our findings suggest that implementing mHealth for weight management requires more than one information session. Findings also suggest that focusing on the coaching relationship and how users' lives and goals change over time may be an important way to facilitate engagement and improved health. Most participants thought mHealth was appropriate for weight management, with some nevertheless preferring in-person care. Therefore, the best way to guarantee equitable care will be to ensure multiple routes to achieving the same behavioral health goals. Veterans Health Administration patients have the option of using mHealth for weight management, but can also attend group weight management programs or single-session nutrition classes or access fitness facilities. Health care policy does not allow such access for most people in the United States; however, expanded access to behavioral weight management is an important long-term goal to ensure health for all.

***(JMIR Form Res 2021;5(12):e29916)*** doi:10.2196/29916

**KEYWORDS**

mHealth; implementation; adoption; engagement; weight management; obesity; weight loss; mobile health; veterans; barriers



XSL•FO
RenderX

## Introduction

Obesity affects roughly 40% of adults in the United States [1] and is associated with many negative outcomes, including social stigma and chronic conditions, such as diabetes and heart disease [2,3]. Although weight management requires long-term, complex behavior change, behavioral weight management programs result in reduced weight, morbidity, and mortality [4-6]. For example, MOVE!, the Veterans Health Administration's (VHA) flagship weight management program, has been associated with reductions in cardiovascular disease [7] and reductions in diabetes complications [8]. However, while 94% of VHA patients with overweight or obesity are offered weight management programs, only 10% use them [9]. Research outside the VHA in the United States does not typically focus on initial engagement in weight management programs because it is difficult to define the denominator of the target population. However, existing work outside Veterans Affairs (VA) also suggests low engagement rates [10-14].

Mobile health (mHealth) interventions for weight management can result in weight loss outcomes comparable to in-person treatments, although results are mixed [15-17]. For example, Track is an mHealth intervention for adult patients with obesity and related comorbidities (eg, diabetes, hypertension) [18]. It provides app-based self-monitoring in addition to dietician-delivered counseling calls with tailored feedback. In a 12-month effectiveness randomized controlled trial at a community health center, participants randomized to Track were more likely to lose at least 5% of their baseline weight. Given the ubiquity of smartphones [19], increasing access to mHealth for weight management could increase weight management program use and improve patient health. However, there is relatively little information on implementing mHealth for weight management in routine practice settings, particularly among older adults and with regard to patient-level factors [20].

In this work, we performed a qualitative analysis of patient interviews to answer two research questions about implementation outcomes among a sample of primary care patients with obesity: (1) what are barriers and facilitators to mHealth adoption and engagement? and (2) what are patient beliefs about the appropriateness of mHealth for weight management? We conducted this work at the VA Palo Alto Health Care System, which had recently implemented an evidence-based [21] commercial smartphone app designed to facilitate weight management and other health behaviors. Starting in March 2019, VA Palo Alto patients were offered access to this app for at least 6 months. The mHealth program includes app-based content (eg, self-monitoring, information modules, and exercises) and a live coach to facilitate goal setting and attainment. Depending on patient preference, the content on the app can be used alone or with coach support. There is also more formal coach-supported content (eg, sessions based on the Diabetes Prevention Program). As noted by Hermes and colleagues [20], obtaining patient-level information on adoption of, engagement with, and appropriateness of mHealth interventions is especially important for implementing patient-facing mHealth. VHA was an important setting for this work as VHA patients represent a population not typically studied in mHealth contexts (eg, older adults [22]).

## Methods

### Recruitment and Interviews

Starting in March 2019, the VA Palo Alto Health Care System offered primary care patients with obesity access to an mHealth program for weight management (Vida [21]). The mHealth program was tailored for VHA patients—for example, there was an effort to hire coaches with firsthand military experience or military experience via family members, or those who worked in the military as civilians. Coaches also received online training in military culture via VHA's Talent Management System.

The program was advertised to and by primary care clinicians, behavioral health staff, and weight management clinicians. It was also advertised at the main hospital's weekly farmers market, and through flyers and social media. In consultation with the Public Affairs Office, ads and images were selected to represent a diverse population with regard to age, sex, and ethnicity. After learning about the mHealth program in one of the aforementioned ways, patients had to contact VA Palo Alto staff, express interest, and meet minimal criteria (ie, access to a smartphone with internet, no uncontrolled mental or physical health conditions). Patients were then given an access code they could use to download the app and begin the mHealth program.

Between February 2020 and October 2020, we used administrative data to obtain a randomly chosen list of 77 potential participants who were VA Palo Alto adult primary care patients with obesity (ie, body mass index ≥30) who had been given an mHealth access code. We excluded 36 patients from this list who no longer had a BMI ≥30, or who had a hospitalization or suicide attempt in the past 30 days, cognitive impairment, and/or a psychotic disorder diagnosis. We mailed opt-out letters to the remaining 41 patients, calling potential participants who did not opt out or otherwise contact study staff. Interested and eligible patients completed the consent process and a one-time interview over the phone (n=24 for a response rate of 59%). We mailed opt-out letters in February, August, and October 2020. We completed interviews in February, March, June, September, October, and November 2020. Therefore, data collection was complete by November 2020. All participants who completed the consent process also completed the interview. Interviews lasted roughly 45-60 minutes. Detailed, typewritten notes were taken during interviews. COVID-19–related technology problems prevented staff from recording all interviews. However, roughly half of the interviews were digitally audio-recorded (n=11). In most cases, each interview consisted of one staff member and one participant. The semistructured interview guide covered beliefs about weight and weight management and was part of a larger study on that topic. Questions most relevant to the present work are in Table 1 (full guide available upon request). We paid participants $50.



XSL•FO
RenderX

**Table 1.** Most relevant interview questions and prompts.

| Question | Relevant prompts |
|---|---|
| Have you ever tried to lose weight? | • N/A[a] |
| Why did you try to lose weight? | • N/A |
| How have you tried to lose weight? | • How/where did you learn about/use Vida?<br>• What made it easy to learn about Vida?<br>• What made it difficult to learn about Vida? |
| What got in the way of you trying to lose weight? | • What got in the way of using Vida?<br>• Would tailored information have changed your experience with Vida?<br>• Did the coach and examples seem relevant to you?<br>• What would have helped you use Vida?<br>• What do you think of the idea of an app with a live coach to help with weight management? |
| What helped you try to lose weight? | • What helped you use Vida?<br>• How did the coach affect your experiences with Vida? |

[a]N/A: not applicable.

### Analysis

We used rapid qualitative analysis methods informed by Neal and colleagues [23] and Nevedal and colleagues [24] to answer the research questions: (1) what are barriers and facilitators to mHealth adoption (initial use) and engagement (continued use, ie, 2 or more uses)? and (2) what are patient beliefs about the appropriateness (ie, perceived fit, relevance, or compatibility [20]) of mHealth for weight management? Rapid analysis has been shown to be faster than and as effective as other forms of qualitative analysis for relatively straightforward research questions such as ours [24,25].

During and after each interview, we completed a matrix in Microsoft Excel (Microsoft Corp) to identify key information about participants' mHealth experiences. The matrix had a row for each participant. Columns included deductive codes related to our primary research questions: whether participants used Vida, mHealth barriers, and mHealth facilitators. There were societal barriers and facilitators columns to note factors related to sociodemographic information and/or discrimination. There was also an "other notes" column, where researchers could document information that did not fit into the aforementioned columns. After all other information was entered into the matrix, we added columns for participants' gender, age, and race/ethnicity to help us identify whether there were patterns across those demographic groups. Analysts also wrote a memo for each participant describing the entire interview.

After all interviews were complete, JYB and KA separately performed inductive analysis by reviewing each barriers and facilitators column and looking for themes—that is, repetition of topics, salience of topics to participants, and negative cases (cases that were different or unique compared to other people). They used the same process across rows to identify possible themes within participants, finding none. The authors then met to review themes and resolve discrepancies (there were few), following which the two authors separately reviewed interview memos to search for additional information on existing themes or information on new themes. After meeting to compare results, noting high agreement and no new themes, the authors settled on a final list of themes and representative quotes from participants. All methods were approved by the Stanford University Institutional Review Board.

## Results

### Overview

We interviewed 24 veterans, 6 of whom were women. Only 2 participants were aged <45 years, and we had roughly equal numbers of participants aged 45-64 years and ≥65 years. Two-thirds of our sample was White; other participants reported their race/ethnicity as Black or African American, Asian, Hispanic, or other. Table 2 provides additional detail. A total of 7 participants reported adopting the mHealth program, 4 of whom were women. For simplicity and to use the language of mHealth, in the sections below we refer to participants who adopted the mHealth program as users and those who did not adopt the mHealth program as nonusers.



XSL•FO
RenderX

**Table 2.** Participant characteristics.

| Characteristics | Total | Adopted mHealth intervention | Did not adopt mHealth intervention |
|---|---|---|---|
| **Gender** | | | |
| Woman | 6 | 4 | 2 |
| Man | 18 | 3 | 15 |
| **Race/ethnicity** | | | |
| White | 16 | 5 | 11 |
| Black or African American, Asian, Hispanic, and/or other | 8 | 2 | 6 |
| Total | 24 | 7 | 17 |

We identified three themes related to the adoption of, engagement with, and appropriateness of mHealth, which are described below. Other than noting a greater proportion of women used the mHealth program than men, we identified no themes related to gender, age, or race/ethnicity. Some findings did not fit under the specific themes, but influenced the presentation of themes. First, most participants had tried multiple weight loss methods, including other forms of mHealth. Second, barriers and facilitators to mHealth adoption were similar between users and nonusers, although the latter necessarily described hypothetical reasons. Finally, improving overall health was the main weight management motivator for participants.

## Theme 1: Coaching Both Facilitates and Prevents mHealth Adoption and Engagement

A participant who used the mHealth program described this theme most succinctly: "the coach…is a double-edged sword, people need accountability, but at the same time they are sometimes afraid of accountability."

Participants said the live coach facilitated adoption and engagement by facilitating accountability. As a user noted, the coach "makes you responsible to answer to someone," adding that the accountability was "a big part of [my] success, setting goals and expectations helped me a lot." This was also true for nonusers, with seven of 17 participants who did not use the mHealth program saying they liked the idea of receiving health coaching via mHealth and two expressing a desire to get another the mHealth program access code after the interview. As one participant who did not use the mHealth program said, the coach was appealing because "I always desire to have someone whom I am accountable to, who is knowledgeable and supportive. In the end we are all humans and need support." Of note, users and nonusers generally did not report a desire to have a coach matched to them solely on demographic characteristics ("I don't care if they're male, female, Black, white"). Instead, they focused on the importance of connecting to the coach in at least one way, which they were as likely to describe as being related to hobbies (eg, types of physical activity) versus a demographic characteristic (eg, gender).

At the same time, participants also described how coaches could prevent adoption and engagement. First, feelings of guilt engendered by not meeting goals led participants to stop meeting with the coach and to stop using the app. For some participants, this was also linked to shame, with one participant stating the following:

> *I didn't like the feeling of answering to someone when I am not successful, I felt like I wasn't doing enough, and I used to feel ashamed when I would think about talking to the coach. This made me resistant to using Vida.*

For this participant, the feelings applied to doctors as well. For others, the guilt was linked to past health struggles:

> *I can't think of any drug addict or person who has a flaw that likes to continuously focus on it. For me it was like, jeez, I have to keep talking about this. Did I do this food? Did I do this activity? And it's like ah, I didn't do it today. I would feel bad because I didn't stick to the plan. I know I'm not disappointing her, but that's just the way it would go.*

Notably, this participant acknowledged that the coach was not judgmental, but that did not alleviate feelings of guilt. Comments from nonusers also suggest a fear of disappointing even a hypothetical coach:

> *Yes, what I really like about the idea is being accountable to somebody. Being able to talk to someone who can understand your issue and is going to be helpful. I would like my coach to be encouraging, supportive and understanding. [A coach that doesn't] make a big deal when you miss something [and] instead says, "it's okay you missed it today it's not end of the world, start a new day from tomorrow."*

High coach turnover rates also prevented mHealth engagement as it resulted in some participants having multiple coaches. Participants described frustration arising from multiple rapport-building attempts with new coaches, and ultimately disengagement. As one participant said:

> *[The coach] enhanced the whole experience, I looked forward to talking to her every week, but then I stopped losing weight which frustrated me…and then [I] had to travel and my counselor changed. I didn't feel like building a relationship with a new person.*

## Theme 2: Preferences Regarding the Mode of Treatment Delivery, Usability, and Treatment Content



XSL•FO
RenderX

### Were Barriers to mHealth Appropriateness and Adoption

While participants said mHealth was generally appropriate for weight management, they also noted that it was not appropriate in all cases. In this way, the mode of treatment delivery served as a barrier for users and nonusers alike. For example, some participants simply preferred in-person treatment. As a nonuser noted, "I like the convenience of phone meetings, but there is a better accountability when you have in person appointments." Another nonuser suggested connecting patients to peers, for example, "through a connected app where you're connected with somebody else in a different part of the country…and you're both going to lose weight together. You have a friend and you're committed with somebody."

Some users described a preference for in-person treatment, despite being willing to try mHealth. One noted that the mHealth program may not have worked for her because the coach "was on the video thing…she wasn't in my face, she didn't know me a whole lot," adding that "it's a physical thing, in person, not over the phone. I just didn't connect with [virtual care]."

Half of participants who adopted the mHealth program thought usability was poor—namely, that entering dietary information was more difficult than other apps they used. Some participants noted they were already using a different kind of mHealth for weight management. With regard to treatment content, several participants said they did not use the mHealth program or did not like using the mHealth program because they did not like monitoring their food. Although, one participant noted that "I can't say that I enjoy logging my food, but I have to acknowledge that it makes a difference."

Participants also noted that treatment preferences could change with life circumstances and competing demands. For example, some noted that they learned about the mHealth program toward the end of MOVE! and thought it would serve as way to maintain weight management behaviors once MOVE! ended. However, for at least one participant, by the time MOVE! ended, his daughter's needs superseded his own goals and prevented mHealth adoption. For a user, her husband's mental health care took a substantial amount of time, which left her limited time to engage in health behavior change. In addition, as another user noted, one's own health could prevent engagement in mHealth-related behaviors: "I was swimming for a few months and then I had an open sore and I [couldn't] get into the pool."

### Theme 3: A Single Invitation Was Not Sufficient to Facilitate Adoption of a New mHealth Program

All participants in this study were given an access code for the mHealth program. However, the most common—and therefore primary—barrier to mHealth adoption among participants was not remembering they received an access code, with 12 of 17 nonusers giving this response. Of these 12, two participants remembered seeing ads for the mHealth program, suggesting some familiarity with the program.

## Discussion

### Principal Findings

This is one of the first studies to assess adoption and appropriateness of mHealth for weight management among a group of middle-aged and older adults in a real-world setting. Most participants viewed mHealth as appropriate for weight management. However, despite seeking out and being given free access to an app with a live coach, the majority did not complete the registration process. As a result, the primary barrier to adoption was not remembering they had access, suggesting potential problems with the feasibility of the implementation of this mHealth intervention. For those who adopted the mHealth intervention, barriers to engagement included guilt and shame related to not meeting goals, competing demands of everyday life, and poor usability. Notably, these barriers were not related to the age of participants. Other than women being more likely to use the mHealth program, we did not find noticeable patterns by gender, age, race, or ethnicity.

Perhaps the most novel finding is that coaching was a double-edged sword for participants. Coaching enticed people to adopt the mHealth intervention. However, if they did not consistently achieve their goals, coaching became a barrier to continued engagement. Both guilt (discomfort with one's actions) and shame (feeling like a person who cannot achieve goals) were forces for disengagement [26]. There is a literature on guilt and shame in relation to health behaviors. For example, Thøgersen-Ntoumani and colleagues found that higher ratings of self-compassion in response to dietary lapses were associated with less guilt and stronger goal perseverance intentions and self-efficacy during weight loss attempts [27]. These results suggest that promoting self-compassion may encourage participation in coaching interventions when guilt and shame may be at play. However, guilt, shame, and self-compassion have not been a focus of the health coaching literature. Our findings suggest this may be an important area of future research. It is possible that mHealth coaching for weight management requires more advanced training to build the rapport necessary to overcome guilt and shame.

In addition, our findings highlight the importance of acknowledging and working with ambivalence during behavior change interventions. Motivational interviewing, which is a counselling method used to resolve ambivalence, is associated with weight loss in the context of eHealth and telehealth weight loss programs [28]. Although it is a part of Vida coach training, our findings suggest that coaches may need to spend additional time addressing ambivalence to ensure continued engagement. Further, our findings related to guilt, shame, and ambivalence may explain why coaching did not improve outcomes in an mHealth intervention designed to increase physical activity among veterans, even though coaching was front-loaded to increase engagement [29]. These factors should be addressed in future research.

It is also notable that participants generally did not have demographic criteria for coaches. Instead, they focused on a need to connect with the coach across varied affinities. In this study, the mHealth program was adjusted and coaches were



XSL•FO
RenderX

trained to account for military culture. Therefore, while not mentioned by participants, it is possible that military-related tailoring was sufficient for participants to feel an affinity to the coach. This is heartening given that the intersectional nature of identity [30] means that it is impossible to demographically match all patients and coaches. At the same time, data from in-person medical settings suggest that demographic matching can influence health outcomes; for example, mortality is lower for Black infants when Black physicians provide care [31]. An important area of future research includes determining whether outcomes are better when coaches are matched on specific characteristics and whether this is more important for historically marginalized populations.

Our results also add to other findings from VHA patients using a web-based weight management program. A qualitative study of women veterans using an online version of the Diabetes Prevention Program had some similar results—the women liked the ability to access intervention materials at any time on the internet, but some did not like self-monitoring [32]. That digital program resulted in higher participation rates than standard VHA weight management programming [33], further highlighting the potential utility of digital health in this context.

Our most actionable finding may be that a single referral to mHealth is insufficient for adoption. Although this finding is related to a relatively simple theme, solutions may be complex. The Expert Recommendations for Implementing Change (ERIC) suggest implementation strategies that might facilitate adoption, including developing and distributing educational materials, obtaining and using patient and family feedback, and using mass media [34]. In the context of mHealth, this could include repeated follow-up contacts via phone, SMS text message–based reminders, and/or marketing materials sent to interested patients more than once. Implementation efforts of other digital health programs in VHA suggest that engaging clinicians and leadership will be especially important [35]. Health care systems could also learn from the private sector while using resources like the Digital Health Checklist to ensure a match between private sector and health care ethical standards [36].

## Limitations

Limitations include a focus on VHA participants. VHA is an integrated health care system that provides primary and specialty physical and mental health care to its patients. People receiving care in stand-alone clinics may have different experiences. In addition, we did not have objective information on whether participants used a specific mHealth program within the app (eg, a program that had a specific number of sessions) as barriers and facilitators could differ across finite versus infinite programs. We also do not have information regarding participants' views of military-related tailoring. Although the random sample of users invited to participate in this research should have helped account for potential differences between people who used the program earlier versus later, given our small sample and limited information about the denominator of people offered the program, the sample may not be representative.

The primary limitation of this study relates to a primary conclusion—few participants who received an access code to a new mHealth program adopted it. Including fewer users than nonusers in our sample could have led to missed themes related to adoption. However, similarities in descriptions of barriers and facilitators to adoption between users and nonusers ameliorate this concern. Other strengths include our older sample and an analysis based on real-world mHealth use, which are not commonly studied.

## Conclusions

Our findings suggest that implementing an mHealth intervention for weight management in an integrated system primarily serving older adults requires more than one information session. Findings also suggest that focusing future research on the coaching relationship and how users' lives and goals change over time will be important for facilitating engagement and improved health. Most participants thought the mHealth intervention was appropriate for weight management, with some nevertheless preferring in-person care. Therefore, the best way to ensure equitable care will be to ensure multiple routes to achieving the same behavioral health goals. VHA patients have the option of using mHealth for weight management, but can also attend group weight management programs or single-session nutrition classes, or access fitness facilities. Health care policy does not allow this for most people in the United States; however, expanded access to behavioral weight management programs is an important long-term goal to ensure health for all.

## Acknowledgments

The authors would like to thank Dr Andrea Nevedal for her methodological insight. The views are those of the authors and do not necessarily represent those of the Department of Veterans Affairs (VA) or the United States government. JYB is supported by a VA Health Services Research & Development Career Development Award (CDA 15-257) at the VA Palo Alto. This work was supported by the Research in Implementation Science for Equity (RISE) program funded by the National Heart, Lung, and Blood Institute (5R25HL126146-05) through the Programs to Increase Diversity Among Individuals Engaged in Health-Related Research (PRIDE).

## Authors' Contributions

JYB conceptualized the study, collected data, conducted analysis, and led writing of the manuscript. KA collected data, conducted analysis, and contributed to writing the manuscript. RM contributed to writing the manuscript.



XSL•FO
RenderX

## Conflicts of Interest

None declared.

## References


1. Ogden CL, Fryar CD, Martin CB, Freedman DS, Carroll MD, Gu Q, et al. Trends in Obesity Prevalence by Race and Hispanic Origin-1999-2000 to 2017-2018. JAMA 2020 Sep 22;324(12):1208-1210 [FREE Full text] [doi: 10.1001/jama.2020.14590] [Medline: 32857101]
2. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. J Am Coll Cardiol 2014 Jul 01;63(25 Pt B):2985-3023 [FREE Full text] [doi: 10.1016/j.jacc.2013.11.004] [Medline: 24239920]
3. Flint SW, Oliver EJ, Copeland RJ. Editorial: Obesity Stigma in Healthcare: Impacts on Policy, Practice, and Patients. Front Psychol 2017 Dec 11;8:2149 [FREE Full text] [doi: 10.3389/fpsyg.2017.02149] [Medline: 29312036]
4. Jackson SL, Long Q, Rhee MK, Olson DE, Tomolo AM, Cunningham SA, et al. Weight loss and incidence of diabetes with the Veterans Health Administration MOVE! lifestyle change programme: an observational study. Lancet Diabetes Endocrinol 2015 Mar;3(3):173-180 [FREE Full text] [doi: 10.1016/S2213-8587(14)70267-0] [Medline: 25652129]
5. Kahwati LC, Lance TX, Jones KR, Kinsinger LS. RE-AIM evaluation of the Veterans Health Administration's MOVE! Weight Management Program. Transl Behav Med 2011 Dec;1(4):551-560 [FREE Full text] [doi: 10.1007/s13142-011-0077-4] [Medline: 24073079]
6. Romanova M, Liang L, Deng ML, Li Z, Heber D. Effectiveness of the MOVE! Multidisciplinary weight loss program for veterans in Los Angeles. Prev Chronic Dis 2013 Jul 03;10:E112 [FREE Full text] [doi: 10.5888/pcd10.120325] [Medline: 23823701]
7. Jackson SL, Safo S, Staimez LR, Long Q, Rhee MK, Cunningham SA, et al. Reduced Cardiovascular Disease Incidence With a National Lifestyle Change Program. Am J Prev Med 2017 Apr;52(4):459-468 [FREE Full text] [doi: 10.1016/j.amepre.2016.10.013] [Medline: 27939239]
8. Jackson SL, Staimez LR, Safo S, Long Q, Rhee MK, Cunningham SA, et al. Participation in a National Lifestyle Change Program is associated with improved diabetes Control outcomes. J Diabetes Complications 2017 Sep;31(9):1430-1436 [FREE Full text] [doi: 10.1016/j.jdiacomp.2017.06.001] [Medline: 28673663]
9. Littman AJ, Damschroder LJ, Verchinina L, Lai Z, Kim HM, Hoerster KD, et al. National evaluation of obesity screening and treatment among veterans with and without mental health disorders. Gen Hosp Psychiatry 2015;37(1):7-13. [doi: 10.1016/j.genhosppsych.2014.11.005] [Medline: 25500194]
10. Ahnis A, Riedl A, Figura A, Steinhagen-Thiessen E, Liebl ME, Klapp BF. Psychological and sociodemographic predictors of premature discontinuation of a 1-year multimodal outpatient weight-reduction program: an attrition analysis. Patient Prefer Adherence 2012;6:165-177 [FREE Full text] [doi: 10.2147/PPA.S28022] [Medline: 22442628]
11. Brace AM, Padilla HM, DeJoy DM, Wilson MG, Vandenberg RJ, Davis M. Applying RE-AIM to the evaluation of FUEL Your Life: a worksite translation of DPP. Health Promot Pract 2015 Jan;16(1):28-35. [doi: 10.1177/1524839914539329] [Medline: 24942749]
12. Brook E, Cohen L, Hakendorf P, Wittert G, Thompson C. Predictors of attendance at an obesity clinic and subsequent weight change. BMC Health Serv Res 2014 Feb 20;14:78 [FREE Full text] [doi: 10.1186/1472-6963-14-78] [Medline: 24552252]
13. Gill R, Karmali S, Hadi G, Al-Adra D, Shi X, Birch D. Predictors of attrition in a multidisciplinary adult weight management clinic. Can J Surg 2012 Aug 1;55(4):239-243 [FREE Full text] [doi: 10.1503/cjs.035710] [Medline: 22617538]
14. Reddy P, Hernan AL, Vanderwood KK, Arave D, Niebylski ML, Harwell TS, et al. Implementation of diabetes prevention programs in rural areas: Montana and south-eastern Australia compared. Aust J Rural Health 2011 Jun;19(3):125-134. [doi: 10.1111/j.1440-1584.2011.01197.x] [Medline: 21605225]
15. Bardus M, Smith JR, Samaha L, Abraham C. Mobile and Web 2.0 interventions for weight management: an overview of review evidence and its methodological quality. Eur J Public Health 2016 Aug;26(4):602-610. [doi: 10.1093/eurpub/ckw090] [Medline: 27335330]
16. Dounavi K, Tsoumani O. Mobile Health Applications in Weight Management: A Systematic Literature Review. Am J Prev Med 2019 Jun;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
17. Schippers M, Adam PCG, Smolenski DJ, Wong HTH, de Wit JBF. A meta-analysis of overall effects of weight loss interventions delivered via mobile phones and effect size differences according to delivery mode, personal contact, and intervention intensity and duration. Obes Rev 2017 Apr;18(4):450-459. [doi: 10.1111/obr.12492] [Medline: 28187246]
18. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an App and Provider Counseling for Obesity Treatment in Primary Care. Am J Prev Med 2018 Dec;55(6):777-786 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]




XSL•FO
RenderX

19. Mobile Fact Sheet. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-10-24]
20. Hermes ED, Lyon AR, Schueller SM, Glass JE. Measuring the Implementation of Behavioral Intervention Technologies: Recharacterization of Established Outcomes. J Med Internet Res 2019 Jan 25;21(1):e11752 [FREE Full text] [doi: 10.2196/11752] [Medline: 30681966]
21. Mao AY, Chen C, Magana C, Caballero Barajas K, Olayiwola JN. A Mobile Phone-Based Health Coaching Intervention for Weight Loss and Blood Pressure Reduction in a National Payer Population: A Retrospective Study. JMIR mHealth uHealth 2017 Jun 08;5(6):e80 [FREE Full text] [doi: 10.2196/mhealth.7591] [Medline: 28596147]
22. Eibner C, Krull H, Brown K, Cefalu M, Mulcahy A, Pollard M. Current and Projected Characteristics and Unique Health Care Needs of the Patient Population Served by the Department of Veterans Affairs. Rand Corporation. Santa Monica, CA; 2015. URL: https://www.rand.org/pubs/research_reports/RR1165z1.html [accessed 2021-10-26]
23. Neal JW, Neal ZP, VanDyke E, Kornbluh M. Expediting the Analysis of Qualitative Data in Evaluation. American Journal of Evaluation 2014 May 27;36(1):118-132. [doi: 10.1177/1098214014536601]
24. Nevedal A, Reardon C, Widerquist M, Jackson G, Cutrona S, White B. Rapid versus traditional qualitative analysis using the Consolidated Framework for Implementation Research (CFIR). Implementation Science 2021;16(1):1-12. [doi: 10.21203/rs.3.rs-137424/v1]
25. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 01;14(1):11 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
26. Miceli M, Castelfranchi C. Reconsidering the Differences Between Shame and Guilt. Eur J Psychol 2018 Aug;14(3):710-733 [FREE Full text] [doi: 10.5964/ejop.v14i3.1564] [Medline: 30263080]
27. Thøgersen-Ntoumani C, Dodos LA, Stenling A, Ntoumanis N. Does self-compassion help to deal with dietary lapses among overweight and obese adults who pursue weight-loss goals? Br J Health Psychol 2021 Sep;26(3):767-788 [FREE Full text] [doi: 10.1111/bjhp.12499] [Medline: 33368932]
28. Patel ML, Wakayama LN, Bass MB, Breland JY. Motivational interviewing in eHealth and telehealth interventions for weight loss: A systematic review. Prev Med 2019 Sep;126:105738. [doi: 10.1016/j.ypmed.2019.05.026] [Medline: 31153917]
29. Damschroder LJ, Buis LR, McCant FA, Kim HM, Evans R, Oddone EZ, et al. Effect of Adding Telephone-Based Brief Coaching to an mHealth App (Stay Strong) for Promoting Physical Activity Among Veterans: Randomized Controlled Trial. J Med Internet Res 2020 Aug 04;22(8):e19216 [FREE Full text] [doi: 10.2196/19216] [Medline: 32687474]
30. Crenshaw K. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. Stanford Law Review 1991 Jul;43(6):1241. [doi: 10.2307/1229039]
31. Greenwood BN, Hardeman RR, Huang L, Sojourner A. Physician-patient racial concordance and disparities in birthing mortality for newborns. Proc Natl Acad Sci U S A 2020 Sep 01;117(35):21194-21200 [FREE Full text] [doi: 10.1073/pnas.1913405117] [Medline: 32817561]
32. Moin T, Ertl K, Schneider J, Vasti E, Makki F, Richardson C, et al. Women veterans' experience with a web-based diabetes prevention program: a qualitative study to inform future practice. J Med Internet Res 2015 May 25;17(5):e127 [FREE Full text] [doi: 10.2196/jmir.4332] [Medline: 26006697]
33. Moin T, Damschroder LJ, AuYoung M, Maciejewski ML, Datta SK, Weinreb JE, et al. Diabetes Prevention Program Translation in the Veterans Health Administration. Am J Prev Med 2017 Jul;53(1):70-77 [FREE Full text] [doi: 10.1016/j.amepre.2016.11.009] [Medline: 28094135]
34. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10(1):21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
35. Damschroder LJ, Reardon CM, AuYoung M, Moin T, Datta SK, Sparks JB, et al. Implementation findings from a hybrid III implementation-effectiveness trial of the Diabetes Prevention Program (DPP) in the Veterans Health Administration (VHA). Implement Sci 2017 Jul 26;12(1):94 [FREE Full text] [doi: 10.1186/s13012-017-0619-3] [Medline: 28747191]
36. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. Transl Behav Med 2020 Oct 08;10(4):1004-1015 [FREE Full text] [doi: 10.1093/tbm/ibz074] [Medline: 31120511]

## Abbreviations

**ERIC:** Expert Recommendations for Implementing Change
**mHealth:** mobile health
**VA:** Veterans Affairs
**VHA:** Veterans Health Administration



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 26.04.21; peer-reviewed by D Goodrich, R Kadri; comments to author 16.06.21; revised version received 24.06.21; accepted 17.09.21; published 08.12.21.*

*Please cite as:*
*Breland JY, Agha K, Mohankumar R*
*Adoption and Appropriateness of mHealth for Weight Management in the Real World: A Qualitative Investigation of Patient Perspectives*
*JMIR Form Res 2021;5(12):e29916*
*URL: https://formative.jmir.org/2021/12/e29916*
*doi:10.2196/29916*
*PMID:34889761*

©Jessica Y Breland, Khizran Agha, Rakshitha Mohankumar. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Creating a Digital Toolkit to Reduce Fatigue and Promote Quality of Life in Multiple Sclerosis: Participatory Design and Usability Study

Sarah Thomas[1*], BSc (Hons), PhD; Andy Pulman[1*], MA, PhD; Huseyin Dogan[2], BSc (Hons), MSc, EngD; Nan Jiang[2], BEng, MSc, PhD; David Passmore[2], BSc (Hons); Keith Pretty[2], BSc (Hons), MBA; Beth Fairbanks[2], BSc (Hons); Angela Davies Smith[3], MSCP, MSc; Peter W Thomas[1], BSc (Hons), CStat, PhD

[1]Bournemouth University Clinical Research Unit, Department of Medical Science & Public Health, Faculty of Health & Social Sciences, Bournemouth University, Bournemouth, United Kingdom

[2]Department of Computing & Informatics, Faculty of Science & Technology, Bournemouth University, Bournemouth, United Kingdom

[3]Bristol & Avon Multiple Sclerosis Centre, Southmead Hospital, North Bristol National Health Service Trust, Bristol, United Kingdom

[*]these authors contributed equally

**Corresponding Author:**
Sarah Thomas, BSc (Hons), PhD
Bournemouth University Clinical Research Unit
Department of Medical Science & Public Health, Faculty of Health & Social Sciences
Bournemouth University
Bournemouth Gateway Building
St. Pauls Lane
Bournemouth, BH8 8GP
United Kingdom
Phone: 44 1202 961953
Email: saraht@bournemouth.ac.uk

## Abstract

**Background:** Fatigue is one of the most common and debilitating symptoms of multiple sclerosis (MS), experienced by more than 80% of people with MS. FACETS (Fatigue: Applying Cognitive Behavioral and Energy Effectiveness Techniques to Lifestyle) is an evidence-based, face-to-face, 6-session group fatigue management program for people with MS. Homework tasks are an integral part of FACETS and are currently undertaken in a paper-based form. Feedback from a consultation undertaken with FACETS attendees and health care professionals with experience in delivering the FACETS program suggested that being able to complete the homework tasks digitally would be desirable, potentially enhancing engagement and adherence and enabling on-the-go access to fit into busy lifestyles. Relative to other long-term conditions, there are few apps specifically for MS and, of those available, many have been developed with little or no input from people with MS.

**Objective:** The purpose of this mixed methods study was to create a digital toolkit comprising the homework tasks (eg, activity diary, goal planner, thought diary) of the FACETS program for people with MS, considering end users' unique requirements throughout the design, build, prototyping, and testing stages.

**Methods:** Phase 1 involved the elicitation of detailed user requirements for the toolkit via 2 focus groups with previous attendees of FACETS (n=3 and n=6) and wireframing. Phase 2 involved supervised usability testing with people with MS (n=11) with iterative prototyping. The usability sessions involved going through test scenarios using the FACETS toolkit on an Android test phone with video capture and concurrent think-aloud followed by completion of the System Usability Scale (SUS) and a semistructured interview collecting feedback about design, content, and functionality.

**Results:** The mean SUS score for the digital toolkit was 74.3 (SD 16.8, 95% CI 63.2-85.6; range 37.5-95), which equates to an adjective rating of good and a B grade (70th-79th percentile range) on the Sauro-Lewis curved grading scale. A number of usability and design issues (such as simplifying overall screen flow to better meet users' needs) and suggestions for improvements (such as using location-based services and displaying personalized information and progress via a central dashboard) were addressed and implemented during the usability testing cycle.



XSL•FO
RenderX

**Conclusions:** This work highlights the importance of the participation of people with MS across the entire development cycle, working to a human-centered design methodology to enable a considered and MS-centered solution to be developed. Continued horizon scanning for emergent technological enhancements will enable us to identify opportunities for further improvements to the FACETS toolkit prior to launch. The toolkit supports self-monitoring and management of fatigue and has potential applicability to other long-term conditions where fatigue is a significant issue.

***(JMIR Form Res 2021;5(12):e19230)*** doi:10.2196/19230

**KEYWORDS**

multiple sclerosis; fatigue; self-management; cognitive behavioral; digital health; mHealth; eHealth; development; participatory design; usability testing

## Introduction

### Background

Multiple sclerosis (MS) is a neurological condition affecting the central nervous system. More than 2.5 million people worldwide have MS with over 130,000 in the United Kingdom [1,2] at an estimated cost to the UK economy of £3.3 to £4.2 billion (US $4.5 to $5.7 billion) per annum [3]. Fatigue is one of the most common and debilitating symptoms of MS [4], experienced by over 80% of people with MS [5], and the main reason for stopping work early [6]. FACETS (Fatigue: Applying Cognitive Behavioral and Energy Effectiveness Techniques to Lifestyle) is a group-based, face-to-face fatigue management program for people with MS developed by members of our team that has been shown to be effective in a national multicenter randomized controlled trial (RCT) funded by the UK MS Society [7-11]. In response to the COVID-19 pandemic, some health care professionals have been delivering the FACETS program virtually (via video conferencing) with initial participant feedback promising.

The program is delivered in 6 weekly sessions [8]. A key component of the FACETS program is the homework tasks that provide an opportunity for participants to try out what they have learned and put it into practice in their daily lives (Table 1).

**Table 1.** Overview of the FACETS[a] program.

| Session no. | Session title | Homework elements |
|---|---|---|
| 1 | What is MS[b]-related fatigue? | Activity and fatigue diary; energy measure |
| 2 | Opening an energy account | Rest, activity, and sleep planner |
| 3 | Budgeting energy and smartening up goals | Goal-setting exercise |
| 4 | Stress response; cognitive behavioral model | Fatigue thought diary |
| 5 | Putting unhelpful thoughts on trial | Thought challenge sheet |
| 6 | Recapping and taking the program forward | Keeping-on-track planner |

[a]FACETS: Fatigue: Applying Cognitive Behavioral and Energy Effectiveness Techniques to Lifestyle.

[b]MS: multiple sclerosis.

Given the favorable results from the national multicenter RCT of FACETS [9-11] and subsequent rollout of the program, team members conducted a consultation on behalf of the MS Society to gather views from stakeholders (people with MS and health care professionals) about potential digital delivery models to enhance reach [12]. Self-guided web-based delivery models for fatigue management have shown promise in MS, although dropout rates have tended to be relatively high [13,14], a common issue encountered in eHealth trials [15]. Findings from the consultation indicated that stakeholders considered an online delivery model of FACETS to be of value [12], although not a direct replacement for the face-to-face version. An online e-learning package was deemed the best way to deliver a minimum viable solution—a product requiring only a limited amount of development time that is implemented with a minimal number of features to provide a basic working model with scope for future expansion and improvement. This has now been launched by the MS Society [16].

During the consultation, it also became apparent there were no high-quality free apps that could support digital completion of the FACETS homework tasks. A separate key recommendation, therefore, was to initiate a project to create a free stand-alone digital toolkit consolidating the structured homework elements of the FACETS program [12]. This project forms the focus of our paper.

Smartphone ownership and use is high in people with MS [17-19]. An app format would have the advantage of permitting on-the-go access and real-time symptom logging and use of reminders, potentially enhancing adherence to the homework tasks [12,20].

> *Everywhere I go, I've got my phone. If I've got a few minutes, I sit and fill it in and if it's fresh and current, I wouldn't fill in paperwork. Even on the course, I'd fill in the paperwork the night before, or the morning I was coming to the class. But if I had it on my phone, I'd be more inclined to fill it in.* [Previous attendee of the face-to-face FACETS program] [12]



XSL•FO
RenderX

The 2018 MS Society/Nuffield Trust data and technology report presents a vision of "personalized, coordinated, and empowering care for people with MS, enabled by effective technology," noting that "Digital transformation and the possibilities it provides have not yet been realised within the care and support people with MS access" [21,22]. A complementary mobile solution enabling the FACETS homework elements to be made interactive and portable aligns closely with recommendations from the MS Society data and technology report and action plan, particularly in the areas of having more control over care and accessible and coordinated care [21-23]. It would help to meet the aims of the UK MS Society research strategy in relation to self-management and implementation [24] and would address the third (fatigue) and fourth (self-management) James Lind Alliance research priorities for MS [25].

In their 2017 systematic review of MS apps available from US app stores, Giunti et al [26] noted there were few apps available for MS relative to other long-term conditions such as cancer and diabetes [27-29]. While there are mobile apps available for fatigue management in MS [30-33], they did not meet our requirements of being free to use. To date, only a fatigue management app for cancer, which draws upon cognitive behavioral principles, has been evaluated in a full-scale RCT [34-36]. While there are separate apps available to support self-management of MS symptoms (including diaries and symptom loggers), these did not align sufficiently with the FACETS homework elements, and most were not free.

### Preliminary Work

Findings from a 2018 systematic search and scoping review by Giunti et al [37] suggested that most MS-specific apps lack features desired by people with MS. They and others have called for greater involvement of people with MS and health care professionals before digital solutions are implemented [37-40], noting the importance of understanding condition-specific factors when designing mobile health (mHealth) apps [37-39]. Developers need to consider the requirements of people with MS and possible MS symptoms (blurry vision, reduced fine motor control, cognitive impairment, and fatigue) throughout the development, prototyping, testing, and implementation of any digital solution [21-23,38,41,42].

Initial requirements for the FACETS toolkit were categorized using the MoSCoW framework [43], a simple method used across business disciplines to enable project teams and stakeholders to define requirements: must have (a necessity for meeting the specified goal), should have (beneficial but not essential for a successful product), could have (desirable but not important), and won't have *this time* (future possibilities but not feasible for immediate implementation). Findings from the consultation [12] informed the initial set of draft requirements, and a stakeholder workshop and interviews with service users were conducted to explore client expectations regarding digitization and begin gathering baseline requirements [44]. An affinity diagram and personas (fictional characters that incorporate composite attributes of target users) were created and used in the early design phases [44]. A card-sorting task (involves ordering, grouping, and naming of objects or concepts) was undertaken with the research team to guide the design of the navigational structure [44]. Paper designs were sketched leading to the creation of low-fidelity wireframes and, subsequently, to the first interactive high-fidelity design prototype.

The toolkit was primarily aimed at complementing the existing FACETS face-to-face program and the MS Society online course. This meant that it needed to remain tethered to the existing materials and the way they are structured to ensure consistency. A secondary consideration was that there might be users downloading the toolkit who had not attended FACETS but who might wish to use elements of the toolkit for recording data relating to their MS and fatigue. A further consideration was that in the future the toolkit might be used to share information with health care providers [45] or expanded to include more content from the FACETS program; the toolkit might also be considered relevant for other long-term conditions [46].

Choosing the type of technology to use for the toolkit was challenging as this is a constantly evolving area. One recommendation from Beatty et al [47] in the context of an evaluation of an intervention for cancer-related distress was that future online programs should be multiplatform to facilitate access across a full range of devices. The FACETS toolkit was initially developed for Android—chosen due to its larger market share and lower anticipated cost of development compared to Apple iOS [48]. Given that a key requirement was access to the toolkit without Wi-Fi and that it would potentially need to align with the MS Society's e-learning course [16] (in addition to the face-to-face program) [7], mobile web solutions were not considered suitable. The toolkit was developed using an agile approach following Google's Material Design guidelines [49] and industrial best practices, with reference to the adapted technology acceptance model [50-52]. Google user interface guidelines [53] were followed when creating and using icons and widgets, and core app quality guidelines [54] were followed to ensure a baseline satisfactory user experience. Coding started in October 2018. Several prototypes were developed that incorporated all MoSCoW requirements identified as must have and should have [43]. Initial versions concentrated on implementing basic operational functionality that could then be expanded upon or altered as required.

## Methods

### Study Design

This was a mixed methods study [55] involving quantitative approaches (MS-specific and demographic questions and the System Usability Scale [SUS] [56]) and qualitative approaches (focus groups, think-aloud protocol, and semistructured interviews). Ethical approval for this research was obtained from Bournemouth University (ref. 17430).

### Participants and Recruitment

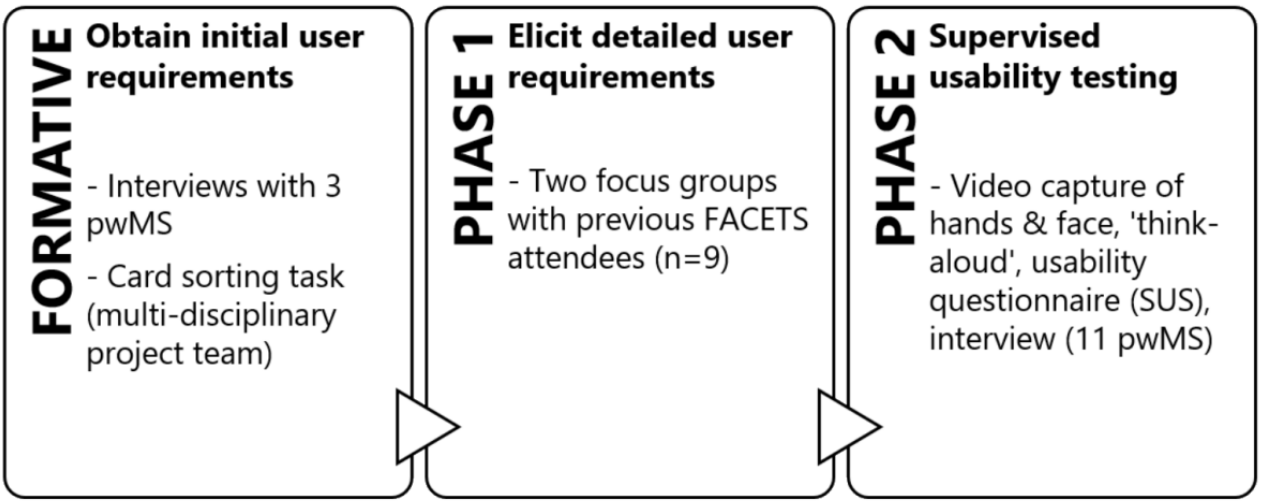
Participants were recruited via a local MS center (individuals who had participated in a previous study and had given permission to be contacted about future research), a local MS support group mailing list, and via an advertisement on the MS Research, Treatment, and Education website. For both the phase



XSL•FO
RenderX

1 focus groups and phase 2 supervised usability testing studies, participants who contacted the research team were emailed participant information sheets with information about the study. Inclusion criteria included being age 18 years or older, having a self-reported diagnosis of MS, experiencing fatigue impacting daily life, and being an active smartphone user (phase 1 only). See Figure 1 for schematic of phases of study.

**Figure 1.** Schematic of study phases. FACETS: Fatigue: Applying Cognitive Behavioral and Energy Effectiveness Techniques to Lifestyle; pwMS: people with multiple sclerosis; SUS: System Usability Scale.




## Procedures

The focus groups (phase 1) were held at a conference center in Bristol. Supervised usability sessions (phase 2) were held at a conference center in Bristol and on the Bournemouth University campus. Both venues were accessible, and taxi-booking and reimbursement of travel expenses were offered to all participants. To minimize participant burden, we obtained written informed consent via a participant agreement form on the day of the focus groups or usability sessions. In phase 2, we audiorecorded and filmed some parts of the supervised usability testing sessions. If participants did not wish to be audiorecorded or filmed, we offered a one-to-one session with notes taken as an alternative. Two copies of the agreement form were countersigned by the researcher. One copy was given to participants for their records, and one copy was retained by the researcher. The main ethical consideration related to fatigue, which is a major issue for people with MS. Focus groups and usability sessions included regular rest breaks and provision of refreshments, and we emphasized that participants could take a break or stop participating at any time. Duration of sessions was no longer than 90 minutes.

### *Phase 1. Focus Groups to Elicit User Requirements*

This phase involved the elicitation of detailed user requirements and wireframing. Two focus groups (n=3 and n=6) with previous attendees of the FACETS face-to-face program were held to gather feedback about requirements and preferences for the toolkit (Multimedia Appendix 1). These were facilitated by AP (in Bristol, an MS assistant practitioner also attended) and audiorecorded and transcribed verbatim. A second set of design prototypes was then created using an interactive prototyping tool, and feedback was obtained from focus group attendees in person and via paper-based semistructured questionnaires completed after attendance. Based on this feedback, a third set of design prototypes was created prior to development commencing.

### *Phase 2. Supervised Usability Testing Sessions*

Participants (n=11) were asked to use elements of the toolkit and complete up to 2 specific test scenarios (lasting 30 minutes in total) on a supplied Android mobile phone (Table 2). Videocapture of their hands (via a usability rig) and face/top half of their body was undertaken as they interacted with the toolkit while concurrently thinking aloud (give a running commentary). They also completed demographic and MS-specific questions and the SUS [56]. Participants were asked about initial impressions of the version of the toolkit they had tested and for feedback on its design, content, and functionality during a subsequent 30-minute (audiorecorded) semistructured interview based on a topic guide (Multimedia Appendix 2). The digital health postdoctoral researcher (AP) led the usability sessions (n=11). The developer (DP) attended all sessions, and ST (research psychologist) attended 9 sessions. Testing commenced with a stable prototype release; later release versions were created based on user feedback and tested with users iteratively (see Figure 2 for example screenshots of a prototype release).



XSL•FO
RenderX

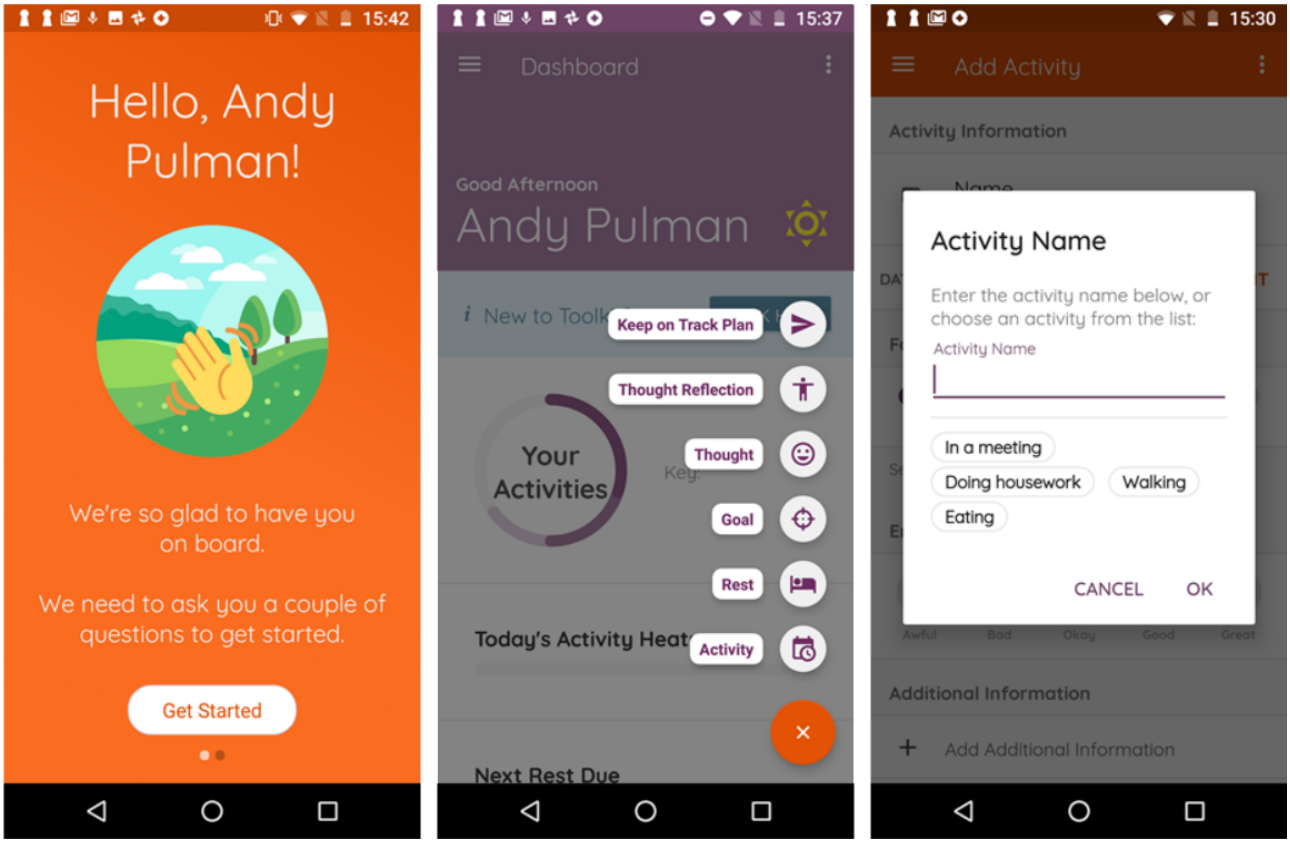
**Table 2.** Mapping of FACETS[a] sessions to the toolkit elements and usability test scenarios.

| Homework task in face-to-face program | Toolkit element | Test scenarios | Participants (1-11) who tested element |
|---|---|---|---|
| Week 1: activity and fatigue diary; energy measure | Activity diary | Add one or more activities (eg, vacuuming, swimming, making breakfast, gardening) performed recently. Make a change to the activity just added, and view it to check that the information was amended correctly. | 1-5 |
| Week 2: rest, activity, and sleep planner | Rest and sleep routine | Thinking about your wake-sleep routine, add a wind-down alarm (for when typically planning to start getting ready for bed) and a wind-up alarm (for when typically planning to start to wake up in the morning). Then create one or more rest periods during the day, setting the duration (scheduling for times when you might wish to take a rest during a typical day). Test out a few alarm options and see if they work correctly. | 1-3 |
| Week 3: goal-setting exercise | Goal planner | Create one or more SMART (specific, measurable, achievable, realistic, time for review) goals (a lifestyle change you would like to make). Some possible ideas for areas for change could be exercise routines, relaxation practice, incorporating rest periods, and establishing sleep-wake routines. | 4-9 |
| Week 4: fatigue thought diary | Thought diary | Think about a situation that triggered strong emotions and unhelpful thoughts related to fatigue. Describe the situation in the thought diary along with up to 3 unhelpful thoughts, and select the emotions experienced at the time. Then rate strength of belief for each unhelpful thought and for the accompanying emotions. | 10-11 |
| Week 5: thought challenge sheet | Thought diary | Select a thought to challenge. Start off by identifying any unhelpful thinking styles, think of one or more alternative thoughts, and rate strength of belief. Then rerate strength of belief in the original thought, and rerate the strength of the associated emotions. | 10-11 |
| Week 6: keeping-on-track planner | Keeping-on-track planner | Help the user maintain momentum: complete a plan that focuses on the next 3 months (this element is still under development). | —[b] (element not yet developed) |

[a]FACETS: Fatigue: Applying Cognitive Behavioral and Energy Effectiveness Techniques to Lifestyle.

[b]Not applicable.

**Figure 2.** Dashboard and homework element screenshots from working prototype v0.0.3.






XSL•FO
RenderX

### Analysis

The SUS [56] is a standardized questionnaire for collecting usability evaluations of a system being tested and has been shown to have good validity and reliability [57], including in the evaluation of mobile health apps [58]. Standard scoring is between 0 and 100 [57], which the Sauro-Lewis curved grading scale [59] converts to a normative percentile score and associated grade. These grades can range from A (best imaginable on the adjective rating scale by Bangor et al [60]) to F (worst imaginable). Participant ratings on the SUS were collected during 2019. Quantitative ratings from the questionnaire were summarized using descriptive statistics.

Focus group and interview recordings were transcribed and thematically analyzed using a deductive approach that focused on the domains covered in the topic guide (focusing on design, functionality, and content). A generic qualitative approach to thematic analysis was used [61] with interresearcher interpretation. Following familiarization with the transcripts, a member of the team charted themes in a matrix. Possible enhancements and amendments to the toolkit (logged by the developer on GitHub) and field notes taken during the think-aloud sessions were also considered in the analysis process. Subsequently, another team member familiarized themselves with the transcripts and the matrix of initial themes. They developed an agreed coding scheme using an analytical framework that combined a priori issues from the original topic guide and emerging themes [62].

## Results

### Participant Characteristics

Focus groups comprised 8 females and 1 male; all participants had previously attended the face-to-face FACETS program. Self-reported descriptives for the usability testing participants (n=11) are presented in Table 3. All study participants consented to being recorded.



XSL•FO
RenderX

**Table 3.** Self-reported descriptives for usability testing participants (phase 2; n=11).

| Characteristic | Value |
|---|---|
| **Gender, n (%)** | |
| Male | 4 (36) |
| Female | 7 (64) |
| Age (years), mean (SD), range | 49 (8.41) 34-62 |
| **Type of MS[a], n (%)** | |
| Relapsing remitting | 8 (73) |
| Secondary progressive | 1 (9) |
| Primary progressive | 2 (18) |
| **APDDS[b], mean (SD), range** | 6.8 (2.5) 2-9[c] |
| 1: mild symptoms that don't limit activity | __[d] |
| 2: noticeable symptoms with mild, small impact | 2 (18) |
| 3: limitations on activities of daily living | — |
| 4: interferes with walking, can walk 300-500 m | — |
| 5: can walk 100-200 m but often uses a stick | — |
| 6: needs a stick or single crutch | — |
| 7: needs 2 canes or walker to walk 20 m | 3 (27) |
| 8: wheelchair is main form of mobility; can move from wheelchair without help | 2 (18) |
| 9: wheelchair is main form of mobility; help needed to move with wheelchair | 2 (18) |
| **Time since diagnosis (years), n (%)** | |
| 1-5 | 3 (27) |
| 6-10 | 2 (18) |
| 11-15 | 3 (27) |
| 16-20 | 2 (18) |
| >20 | 1 (9) |
| **Employment status, n (%)** | |
| Working full time (>30 hours per week) | 4 (36) |
| Unable to work | 3 (27) |
| Retired | 4 (36) |
| **Use of mobile apps, n (%)** | |
| Never use | 1 (9) |
| Use a few | 4 (36) |
| Use a lot | 6 (55) |
| **Has attended face-to-face FACETS[e] program, n (%)** | |
| Yes | 7 (64) |
| No | 4 (36) |
| **SUS[f]** | |
| Median (range) | 75 (37.5-95) |
| Mean (SD) [95% CI] | 74.3 (16.81) [63.2, 85.6] |
| **Sauro-Lewis adjective rating, n (%)** | |
| A/A+ | 5 (45) |
| B | 1 (9) |



XSL•FO
RenderX

| Characteristic | Value |
|---|---|
| C | 2 (18) |
| D | 2 (18) |
| E | 1 (9) |

[a]MS: multiple sclerosis.

[b]APDDS: Adapted Patient-Determined Disease Steps.

[c]Possible scores on the APDDS scale range from 0-10 corresponding to 11 ordinal levels of functioning. However, 2 participants gave ratings indicating they perceived their functioning to fall between ordinals (1 participant between 7-8 and another participant between 8-9) and for the summary statistics these were scored as 7.5 and 8.5, respectively.

[d]Not applicable.

[e]FACETS: Fatigue: Applying Cognitive Behavioral and Energy Effectiveness Techniques to Lifestyle.

[f]SUS: System Usability Scale.

## Focus Groups

Below we summarize the key toolkit requirements suggested by focus group participants:

- Should be a "tool to help rather than a time-consuming task"
- Suitable for those unfamiliar with FACETS
- Important to include positive aspects
- Include self-monitoring feedback (eg, dashboard and graphs)

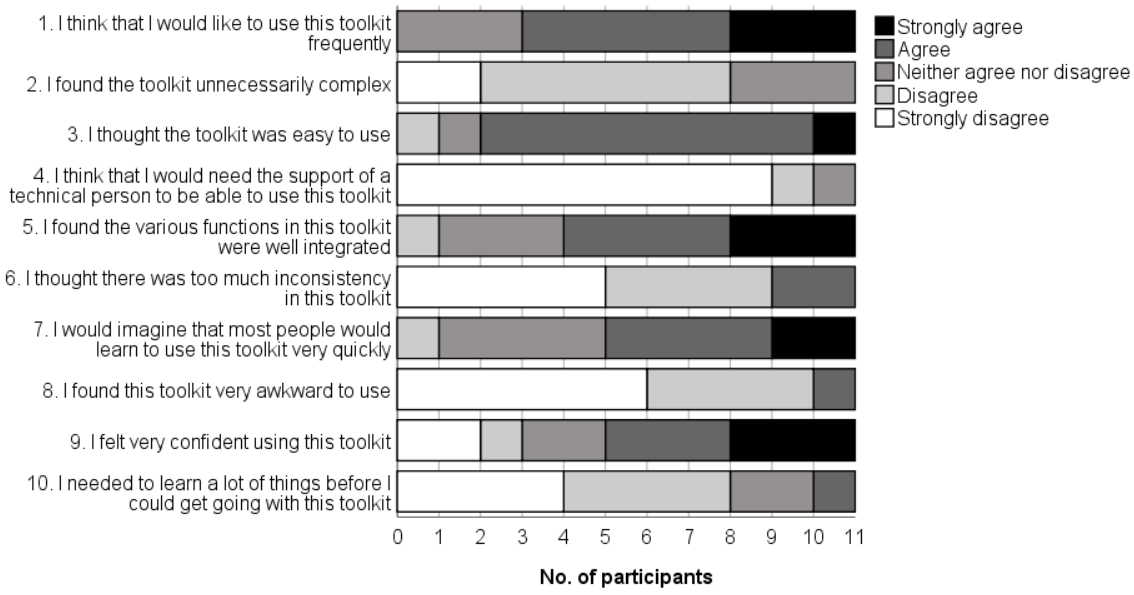
Below we summarize preferences for the toolkit expressed by focus group participants:

- App should be freely available at app stores
- Reminder (eg, reminders to take rests) and note functions would be useful
- Would like relaxation module to be included
- Linking with Alexa is a good idea

## System Usability Scale

Summary statistics for the SUS scores are presented in Table 3, and frequencies of ratings for each of the 10 SUS items are presented in Figure 3. Overall, the median SUS score for the toolkit was 75 and the mean was 74.3 (SD 16.81; 95% CI 63.2 to 85.6; range 37.5-95.0). This equates to an adjective rating of good [60] and a B (70th to 79th percentile range) on the Sauro-Lewis curved grading scale [59]. The majority of participants (8/11) thought they "would like to use this toolkit frequently" (the version of the prototype they tested; SUS-Q1). Most considered the toolkit easy to use (9/11; SUS-Q3) with 55% (6/11) believing most people would learn to use it quickly (SUS-Q7). The majority (7/11, 64%) considered the toolkit's functions to be well integrated (SUS-Q5).

**Figure 3.** System Usability Scale—item response frequencies.




## Usability Feedback

Below we organized the feedback from the usability sessions encompassing general comments and those related to design, functionality, and content, mapped to the toolkit elements. We considered suggestions made for additional functionality. We also included relevant feedback elicited during the focus groups.



XSL•FO
RenderX

### General Comments

Overall, users found the app relatively easy to navigate, liked its look and feel, and felt that they would quickly get used to the system.

> *I did like it and I liked the colors because they are bright but they're not too bright, it was quite, it looked quite vivid I think it is fair to say, which is good. I think it was quite easy to navigate round, I mean once you have got the basic, you know, the three dots and the three lines, from there you can pretty much go anywhere and yes I think it was quite good* [P203, Usability sessions]

> *I found my way around easy. I think initially I just had to get to grips with it. Just to look where things were but...I like the menu structure, it was easy to follow.* [P204, Usability sessions]

However, sometimes users noted difficulty in keeping track of where they were in the app, and early versions of the recap section were found to be confusing.

> *I'm quite visual and I couldn't picture the structure of where I was in the...you know...hierarchy of areas, kind of thing, so that was kind of confusing because I then couldn't then think "Oh I need to skip that bit."* [P202, Usability sessions]

Users wanted content and information to be kept brief and easy to understand, without unnecessary repetition and with options to skip content. As one user noted in a focus group in phase 1, it should be "a tool to help rather than a time-consuming task." Users were in favor of the inclusion of video in addition to audio and text.

> *You have to be careful not to over...put so much information that it becomes overwhelming, that you can analyze down to so much and you think I can't think about this anymore and you put too much into it.* [P211, Usability sessions]

> *I don't want to be spending too much time doing it so I'm missing out on doing things...* [P205, Usability sessions]

> *Yes you know I am ok reading text but I don't want to read too much, so you know it is balancing it between video and text.* [P207, Usability sessions]

Users liked the concept of section-based onboarding and said that prompts and examples were useful, particularly for those with fatigue and cognitive issues. They noted the importance of ensuring reminders are neutral and nonjudgmental in tone.

> *Yes and you know people with MS we do sometimes get flustered and forget how to do things, so a reminder every now and again, the option to be able to go back and actually see how I do this.* [P204, Usability sessions]

> *Yes I think...if somebody is struggling a bit and it keeps saying, "You should be doing this." But if it said, "Would you like to go?" that, that seems a bit more neutral.* [P208, Usability sessions]

One user said they would like the app to "look more like a game" (P2011, Usability sessions). The app was seen to be a helpful self-monitoring tool that would be useful for describing fatigue levels, activity patterns, and symptoms to an individual's clinical team. It was noted that it could also include content that would help family and friends understand more about MS fatigue and its impact.

> *As well as having the ability to be able to explain to someone who doesn't have MS, "Do you know what? Watch this video on this app that I've got."* [P203, Usability sessions]

Users felt that the app could be a useful tool not only alongside the FACETS program but also before or after the program and could be used to gather information prior to a referral by a clinical team for fatigue.

> *But anyway, it would have been really, really good to have had an app with videos or animations [when newly diagnosed] to talk through the premise of FACETS, almost maybe as a precursor to attending the course.* [P203, Usability sessions]

> *Yeah, because my stuff at the moment goes into spreadsheets and when I have to phone up to say, "Oh, I think I am having a relapse," they ask you a whole bunch of questions: "Have you done this?" "Have you done that?" "When was the last episode you had?" "What is your fatigue like?" All these questions it's all on a little Excel sheet on my phone at the moment but this will be a bit more of a proper user interface to get that kind of stuff....* [P204, Usability sessions]

> *I think for me this is a really hugely useful parallel tool alongside the course or, kind of, after the course* [P202, Usability sessions]

> *I mean it would be brilliant for me, I went on the FACETS course a really long time ago, or it feels like a really long time ago now, so to have all of that information would be really good to refresh my memory.* [P203, Usability sessions]

Users noted that the app could be useful for those with other conditions where fatigue is a symptom.

> *I think it could be useful...I think it could be infinitely useful for all people with MS, but not just MS, all people who have fatigue, who have a condition that is...causes fatigue.* [P208, Usability sessions]

### Activity Diary

#### Feedback

The activity diary was seen to be a helpful way of enabling users to consider activity and fatigue patterns.

> *Okay, so yes there, that's quite good for monitoring because actually then you could start to build up a pattern of what makes you feel really fatigued, couldn't you?* [P201, Usability sessions]

> *So it's as if you can put something in so you don't overbook yourself you know...and my thing is that I*



XSL•FO
RenderX

> *am supposed to have two rest days a week you know, it's like reminding people it's fine, rest is part of management really.* [P209, Usability sessions]

Enabling the user to choose from a prepopulated activity list (and eventually using predictive text suggestions) rather than manual entry was viewed as a useful time-saving feature. It would also make a breakdown of activity type possible.

> *I think... I think more is probably, I know it's difficult to put in more, but in terms of fatigue, it's often the quite nuanced type of activity that is worse or better... it wouldn't be particularly helpful to have the broad categories because actually work means doing so many different types of things that um...* [P202, Usability sessions]

The use of location-based services to pull and record relevant information, such as the outside temperature (heat often adversely affects MS fatigue), was suggested by a focus group participant.

> *P101: ...if it can be linked into the weather, like today it is humid so we have got to think take the shade, keep cool.* [Focus Group 1]

#### Prototype Changes Made to This Section

Design changes implemented based on usability feedback included implementing swipe to refresh layouts, updating the calendar view originally used to reduce background thread load, fixing a problem whereby activities entered between 12:00 AM and 1:00 AM were not displaying, and revising the layout to improve flow. Location-based suggestions and the addition of predictive text functionality were implemented.

### *Rest and Sleep Planner*

#### Feedback

Functionality issues reported included the variable quality of some of the alarm tone sounds included as default on the test phone.

> *Participant: ...and I wake up to the sound of birds, which is nice, it's not...*
>
> *Interviewer 1: Yeah, I was looking at those and they are....they are much more...*
>
> *Participant: It is nicer than a belting alarm...* [P202, Usability sessions]

Concern was also expressed by users over the initial layout and legibility of the clock display and rest, wind-up, and wind-down periods chosen (Table 2).

> *Yeah so if I'm looking at that, right ok, especially with my eyesight problems, I've got to read through this whole list to find the thing I want. You've got icons next to it, so sleep and wake has got a little cog and half-moon icon next to it but if that...what I mean is if that was bigger and maybe colorful then I could go to it quicker and say right that's what I want to go back and have a look at my sleep patterns.* [P210, Usability sessions]

#### Prototype Changes Made to This Section

We amended the visuals to include the duration of wind-up and wind-down periods (Table 2). We also ensured it was easy to read by making the layout tabular in order to separate new rest and sleep/wake routines. Particular attention was paid to customizing how the alarm options could be set and configured. Technical problems that required resolution included issues with the ringtone preview and selection on certain devices, and the alarms not resetting upon phone restart.

### *Goal Planner*

#### Feedback

Participants considered the opportunity to set goals a very useful part of the face-to-face program. Suggestions were made to improve user understanding within a screen-based format by offering the ability to view hypothetical worked examples to help convey the concept of SMART (specific, measurable, achievable, and realistic with time for review) goals.

> *...if you click, 'yes I understand that', then yes it goes on to the next bit or if 'no' then the health care professional would tell you what SMART means because you don't need another MS person telling you that.* [P207, Usability sessions]

Users noted that having the ability to select personalized, customizable reminders for entered goals and review and update them could increase engagement with the toolkit.

#### Prototype Changes Made to This Section

The option to note a goal quickly and come back to it later to apply SMART criteria was implemented following user feedback. Case study examples were added to enhance understanding of the concept of SMART goals as this can be difficult for some to understand. Other changes made to improve functionality included adding in pop-up logic to provide guidance to the user when 'no' was selected in the 'realistic' field of the SMART criteria and revising the layout to make it more consistent with the activity diary.

### *Thought Diary and Thought Challenge*

#### Feedback

As can be seen in Table 2, elements 4 and 5 of the FACETS homework had been integrated into one section (Thought Diary) to reduce the complexity of screen flow for the user. Feedback from users suggested that although this section was useful, further simplification was required.

> *Yes I think this thought summary that's useful because, you had to think about how you were thinking in reflection to the impact it's had on you, and I found that quite useful to think, because you do exhaust yourself worrying about everything.* [P211, Usability sessions]
>
> *Well, because I'd already rated it once, that was the original frustration of what that incident or whatever was causing me. I don't think, by the time I'd got to reflecting on the thing and gone back and reviewed what was going on and how I could approach it differently or this, that and the other, I didn't think it*



XSL•FO
RenderX

*made me any less frustrated or angry from the point that I'd already rated it.* [P210, Usability sessions]

**Prototype Changes Made to This Section**

The screen flow and structure was simplified further (eg, by removing the requirement to provide ratings for strength of belief in alternative thoughts).

### *Additional Features and Other Issues*

Adding a notes function (to enable the quick addition of general pieces of information or thoughts for later expansion or reflection) was suggested as a way of reminding users about historical health-related information (eg, about new symptoms that had emerged, ongoing concerns, other scheduled appointments) that could be useful at MS review consultations (sometimes held a year apart).

*When you are newly diagnosed you just, there is this barrage of questions and they don't come to you when you are sat in front of the consultant or even the nurse and you go home and you think, "Oh god I wish I had asked that." Oh do you know what I've just had a thought, maybe having a notes app, a notes section to write down, you know, notes about your [... ] fatigue or notes about anything else MS-related that you need to talk to your nurse or doctor about, that might be nice.* [P203, Usability sessions]

*I use the notes in my phone a lot, so, er, information that I am told...almost everything, if it doesn't go into my phone at the time, it gets lost. For instance, I have got the swimming timetable, which I would then put into my fatigue management diary...* [P208, Usability sessions]

Displaying progress on the dashboard and allowing the user to apply custom queries to data entered (such as their most fatiguing or enjoyable activities) were suggested as ways to enhance engagement. A customizable dashboard could allow the user to create a display containing summary information most relevant and useful to them.

*Yes because I would refer back to this and kind of look at what I have done and gauge my fatigue levels.* [P204, Usability sessions]

*It would be good to have one that was a weekly... so you could see what days you were particularly, you know, it's like actually I need...I have got a lot that I am trying to achieve on that day, so I need to know that I need to have a rest on that day or before....* [P207, Usability sessions]

Although most participants liked the idea of being able to visualize their fatigue levels over time, one participant felt that this could be disheartening. A customizable dashboard would mean that this feedback could be turned off if a user did not find it helpful.

*Yeah if you could turn it on and off or something so you wouldn't have to see it if you didn't want to see it or something, yeah it could, if you had that over several weeks and you're just looking at red, it could be sort of like a bit demoralizing yes.* [P205, Usability sessions]

Users liked the text-to-speech feature that we included in the recap content.

*The fact that you watch or even [with] text you can have somebody saying it at the same time, I find that really helpful for myself, because just reading stuff sometimes it doesn't, it doesn't go in properly.* [P204, Usability sessions]

*It helps there was a voice saying the words as well because it breaks it up.* [P204, Usability sessions]

*The text was really easy to read, but having it read to me was really useful.* [P208, Usability sessions]

Feedback from users in the initial usability sessions suggested they found the synthesized text-to-speech voice overly robotic.

*Then as you went on she turned more robotic and more... yes, less human and I think it would be better to be told by somebody sounding a bit more personal. Whether it's a man or a woman it doesn't really matter I suppose, it's just yeah the right person, but yeah.* [P205, Usability sessions]

As the voice was device-dependent, we only had a certain degree of control over it, but we did find slowing it down a little made it sound more natural.

Usability testers considered the possible integration of voice-activated speakers as a positive method of engagement potentially saving time and energy and requiring less dexterity. In some of the sessions, we gave a demo of inputting a diary entry using Google Assistant, and this was well received.

*Yeah. That's exactly what I would like. Something like that I would use.* [P210, Usability sessions]

Ensuring that help and support functions on the toolkit accommodated the needs of those unfamiliar with the FACETS program was seen as important. Different types and levels of information would be required for those using the toolkit in conjunction with the FACETS face-to-face program or the e-learning course versus those using it independently. The need for this information to be structured more logically to improve the user journey was highlighted in early usability sessions where users experienced difficulties navigating the recap menu.

*So yeah, if I clicked on the activity diary and it says recap, I think I don't know what that is.* [P202, Usability sessions]

## Additional Material Incorporated Into Toolkit

We implemented a notes section as suggested, text-to-speech functionality (available in recap sections) was added, and expanded help sections were developed. The concept of the personalized, customizable dashboard was developed gradually as more toolkit elements were introduced and tested over time.



XSL•FO
RenderX

## Discussion

### Principal Findings

In its current form the FACETS toolkit was evaluated as good on the SUS, and qualitative feedback from usability sessions indicated that users felt the toolkit would be useful. Users provided numerous suggestions for improving the toolkit in terms of design, functionality, and content. Some suggestions were implemented immediately following feedback from usability testers. Although the digital format necessitated the simplification of some aspects of the toolkit elements (to reduce cognitive demands and fatigue), it also presented opportunities to create synergies and interactivity between toolkit elements and the dashboard along with visualization and customization possibilities [63].

Findings from previous research suggest that most MS-specific apps lack features desired by people with MS [37,64], resulting in poor uptake [63]. Giunti et al [37] and others have noted the importance of taking into consideration condition-specific factors when designing mHealth apps [39]. The involvement of people with MS throughout the design, prototyping, and usability phases of the toolkit means that such considerations have played a pivotal role. Examples include providing a customizable color scheme (for those with visual difficulties), using icons where possible (to aid memory), providing signposting, and incorporating guidance (eg, buttons to indicate additional scrolled content or the availability of additional information or definitions).

There was some degree of tension between maintaining consistency with the original paper-based FACETS homework tasks and capitalizing on the possibilities afforded by the digital format. The toolkit was developed based on the homework elements of FACETS but did not provide an exact one-to-one mapping. For example, after review of the initial prototype by the team, the thought diary and thought challenge homework elements were combined into one tool as it was felt that the original screen flow was complex to navigate and could lead to frustration or disengagement by users (Table 2) [65].

### Comparison With Prior Work

As noted in the introduction, there are currently few mobile apps for fatigue management in MS. A review of mHealth in MS by Gromisch et al [66] identified 3 mHealth-based apps and 1 web-based platform that promote fatigue self-management through a variety of approaches [66]. These include cognitive behavior therapy principles (MS Energise, which draws upon the FACETS program [30-32]), gamification of energy management via stamina credits (More Stamina [67,68]), and use of validated self-assessments and medication and activity diaries (MSMonitor; web-based platform) [69]. One app (MS Telecoach) focuses on increasing physical activity levels via telemonitoring (accelerometers and self-reported fatigue) and telecoaching (advice, motivational messages, and goal setting) [70].

Similar to findings from the MS Energise [32] and More Stamina [67] usability studies, user feedback for the digital toolkit suggested a need to simplify some aspects. A cross-national qualitative study on facilitators and barriers to using mHealth tools for managing MS highlighted the importance of clear, simple design and features to enhance user accessibility and engagement [40].

Users in the MS Energise usability study suggested text-to-speech functionality would be helpful [32]. We obtained similar feedback in our early usability sessions and incorporated text-to-speech functionality into the recap sections. Users reported finding this useful, describing how it broke up content and aided concentration.

The More Stamina app incorporates gamified elements in the form of stamina credits [67]. In this study, one user said that they would like the digital toolkit to be more game-like. In previous work, we obtained mixed feedback in relation to gamification from both people with MS and health care professionals in the context of fatigue management [12]. Giunti et al [67] found that in their formative work for More Stamina, people with MS reported a preference for collaborative gamified tasks rather than competitive tasks. Untire, a fatigue management app for cancer-related fatigue, incorporates gamified elements (such as progression bars, rewards, and badges) [34]. Gamified elements could be incorporated as optional features in a future version of the digital toolkit. This will be an important area to explore further with users.

The Untire app for cancer-related fatigue allows users to invite a buddy so that they can manage their fatigue with a family member or friend [34]. Feedback from users in this study suggested that the digital toolkit could be a useful tool to help family members and friends understand more about MS fatigue.

In the MS Energise usability study, users felt the app could be particularly helpful for people with MS soon after diagnosis [32]. Similar comments were made in relation to the digital toolkit, and it was seen to offer potential as a tool that could support communication with the clinical team about fatigue and ongoing monitoring, assessment, and treatment. Most testers in this study had been diagnosed for more than 5 years, so further testing of the toolkit is needed with participants who are relatively newly diagnosed.

During field testing of MS Energise, the authors reported that a task involving the identification of unhelpful thoughts in a fictional character gave rise to unhelpful thoughts in some users about their own fatigue [30]. In our study, one user reported that seeing a visual display or heatmap of their own fatigue ratings could be demoralizing. Findings such as these underscore the importance of working closely with users throughout the development lifecycle.

### Prototyping

Due to the power and graphical capabilities of smartphones, mobile apps can be overdesigned [42,71] leading to the paradox of choice [65], where users struggle to process an overwhelming amount of content or too many interface options. The use of online interactive prototyping tools for initial design iterations helped the project team envision their ideas more effectively than with paper alone. They also helped address some of the challenges of working in different locations and across digital and health fields. We recommend the use of an interactive



XSL•FO
RenderX

prototyping tool to maximize the advantages of participatory design [72,73] in early design stages. Like Giunti et al [38], we found the use of personas (fictional characters that incorporate composite attributes of target users) a helpful way to capture and convey the varied nature of MS during preliminary stages of development [44].

### Mobile Platforms

A mobile platform offers scope for greater personalization than the paper-based FACETS program materials. As noted, while most participants liked the idea of being able to visualize their fatigue levels over time, one participant noted that this could be disheartening. While self-monitoring can help people with long-term conditions to feel more in control [74], it can also evoke negative emotional reactions [75,76]. In the longer term, we intend to offer customization of the dashboard enabling users to specify the information they wish to visualize or focus on so that it is relevant and meaningful to them. We also explored the best ways to support users to make sense of their self-monitoring data [76].

Suggested methods of highlighting progress included using tracking to document accessed sections and displaying progress on a central dashboard, also highlighted in our previous consultation [12]. Over a longer period of time, this might provide the user (and potentially their health care team) with the ability to track data relating to fatigue and could help to reinforce the FACETS program principles and support users in making lifestyle changes. In the longer term, the toolkit offers possibilities for longitudinal symptom and self-monitoring, data sharing, and greater integration of self-management strategies into daily life, with potential applicability to other long-term conditions. For example, fatigue is a significant and debilitating symptom in a number of neuromuscular disorders [77] and post-COVID-19 [78]. The FACETS program has recently been piloted with a group of long COVID-19 patients using a videoconference delivery format [79].

### Digital Tools

Digital tools like wearables and apps are now starting to help users manage the logistics of their long-term conditions, reminding them to take their medication [80] or helping them manage injection sites [81]. In some conditions like diabetes, significant progress has been made in developing and embedding specific technologies to support self-management [82] and empower patients [22], such as flash glucose monitoring [83]. These have yet to make much impact in MS to date [21-23,64], and concerns over the quality, sustainability, and effectiveness of some solutions remain [84]. Global app stores currently have no direct links with official medical organizations like the NHS or obligations to regulate apps on their behalf. This presents potential risks in terms of inaccurate information being made available and subsequently used within a health context [85,86]. Providing hypothetical worked-through examples within the toolkit elements provided a way of conveying key concepts and illustrating potential benefits from their use. In terms of sustainability, due to the number of devices supporting Android, fragmentation within the current market is huge. Most Android devices are currently using versions of the operating system that are more than 2 years old. Considered targeting of a significant portion of the user base is recommended (rather than trying to target every possible Android legacy device). This is especially important given that new versions of Android and Apple operating systems are released yearly to encourage consumers to upgrade their phones regularly.

The process of developing the digital toolkit was highly iterative and agile; where possible we implemented user suggestions for improvements as we went along. For example, in response to user suggestions, we incorporated the capability for real-time weather data in the activity diary (that could be time stamped at time of input or called back retrospectively) and text-to-speech functionality (available in recap sections).

### Emergent Technologies and Future Enhancements

Emergent technological enhancements (such as Google Voice and Google Assistant) offer further opportunities for improved personalized eHealth solutions [87] and increased engagement [88] and adherence [89]. We are currently exploring the use of voice-activated speakers and assistants such as Google Assistant or Amazon Alexa to enable the input of information via voice rather than keyboard, an innovative way to reduce screen fatigue. We have successfully prototyped this in a closed test with further development planned.

Enhancements could also encompass how a digital toolkit could be used intelligently for smart home-based monitoring and assessment via mobile devices [90]. Emerging technologies such as artificial intelligence, machine learning, and remote monitoring of condition markers could provide opportunities for more tailored and personalized care, from services to treatments [21]. For example, personalized advice and support could be offered based on user responses (in the longer term learned by artificial intelligence) to improve engagement.

Areas for further exploration might focus on the use of the toolkit for self-management and monitoring [64] and facilitating communication with health care or other service providers (such as long-term disability benefits assessors [91]) [39,75]. User suggested enhancements made to the prototype included the ability to add notes, which could be used for meetings with health care providers. Other enhancements might include integration of the toolkit with existing data streams. This could include the ability to collect biometric real-time data by using plug-in oximeters and wearable monitoring devices [92] (although improvements in accuracy are required). There is also the possibility of linking recorded data from the toolkit (such as FACETS attendance, activities logged, and goals set) to secure online sources such as the UK MS Register [93]. The toolkit's database design and security structures allow for this possibility.

### Strengths and Limitations

A limitation of this research is that the prototype toolkit has only been tested to date in a closed, controlled environment with a limited number of potential users (n=11). It has been argued, however, that 10 or more testers is adequate to identify the majority of usability issues [94]. A further limitation is that currently the toolkit is only available on Android.



XSL•FO
RenderX

Strengths of this study include the multidisciplinary project team, involvement of people with a range of MS types and mobility throughout, the use of mixed methods, and our agile approach to development. Our testing protocols were designed to take fatigue-related issues into account including ensuring testing locations were fully accessible, providing taxis (if required) to the usability testing location, providing refreshments, and incorporating rest breaks.

### Future Research

The keeping-on-track planner (section 5 of the toolkit) is currently still under development. The next phase of the study will involve people with MS remotely testing the prototype using their own Android smartphones and providing feedback via an online semistructured questionnaire (telephone interviews with a subsample). This will enable additional comments, ideas, and development glitches to be identified by a wider range of users across a broader range of mobile devices and actioned prior to release. We will next pilot the toolkit in conjunction with the FACETS program (either face-to-face or virtual) and the e-learning course (similar to the real-world usability testing of MindClimb, an app developed to support skills practice alongside group cognitive behavioral therapy for anxiety in adolescents [20]).

To date, involvement of health care professionals as stakeholders in the development process has been limited. We plan to gather feedback about the toolkit from health care professionals in our next phase of testing. Development of an iOS version has begun, with the toolkit currently in a pre-alpha stage, and this development will continue. Future plans include expanding the app to include more content from the face-to-face program to provide further scaffolding for the tools in the toolkit and self-management support. As the 2018 MS Society/Nuffield Trust data and technology report suggests [21], rather than immediately reaching for data and technology solutions, the future should be constructed within an environment that enables the potential of data and technology to give people with MS the best care possible.

## Conclusions

We have described a mixed methods approach to the design, prototyping, and usability testing of a digital toolkit comprising the homework tasks of FACETS. This work highlights the importance of the participation of people with MS in the development cycle, working to a human-centered design methodology. Continued horizon scanning for emergent technological enhancements will enable us to identify opportunities for further improvements prior to launch.

## Acknowledgments

The authors would like to thank all participants contributing to the participatory design sessions and the MS Research, Treatment, and Education charity for supporting recruitment. This research forms part of a Higher Education and Innovation Funding project.

## Authors' Contributions

ST and PWT conceived the study. All authors contributed to the study design. BF and AP performed the data collection and requirements generation in phase 1; AP did the same in phase 2. AP and ST analyzed and interpreted the data and conducted qualitative data analysis. ST and AP led the manuscript preparation. All authors contributed to the manuscript preparation, participated in final approval, and take responsibility for the integrity and accuracy of the data analysis.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
User requirements topic guide (phase 1).
[DOCX File , 24 KB - formative_v5i12e19230_app1.docx ]

Multimedia Appendix 2
Interview topic guide (phase 2).
[DOCX File , 29 KB - formative_v5i12e19230_app2.docx ]

## References

1. Tinelli M, Kanavos P, Efthymiadou O, Visintin E, Grimaccia F, Mossman J. Using IMPrESS to guide policy change in multiple sclerosis. Mult Scler 2018 Aug 24;24(9):1251-1255. [doi: 10.1177/1352458517737388] [Medline: 29064774]
2. Multiple sclerosis: prevalence, incidence and smoking status—data briefing.: Public Health England; 2020 Feb 04. URL: https://www.gov.uk/government/publications/multiple-sclerosis-prevalence-incidence-and-smoking-status/multiple-sclerosis-prevalence-incidence-and-smoking-status-data-briefing [accessed 2021-10-05]
3. Kobelt G, Thompson A, Berg J, Gannedahl M, Eriksson J, MSCOI Study Group, European Multiple Sclerosis Platform. New insights into the burden and costs of multiple sclerosis in Europe. Mult Scler 2017 Jul;23(8):1123-1136 [FREE Full text] [doi: 10.1177/1352458517694432] [Medline: 28273775]



XSL•FO
RenderX

4. Desborough J, Brunoro C, Parkinson A, Chisholm K, Elisha M, Drew J, et al. It struck at the heart of who I thought I was: a meta-synthesis of the qualitative literature examining the experiences of people with multiple sclerosis. Health Expect 2020 Oct;23(5):1007-1027 [FREE Full text] [doi: 10.1111/hex.13093] [Medline: 32578287]
5. Zajicek JP, Ingram WM, Vickery J, Creanor S, Wright DE, Hobart JC. Patient-orientated longitudinal study of multiple sclerosis in south west England (The South West Impact of Multiple Sclerosis Project, SWIMS) 1: protocol and baseline characteristics of cohort. BMC Neurol 2010 Oct 07;10(1):88 [FREE Full text] [doi: 10.1186/1471-2377-10-88] [Medline: 20929556]
6. Raggi A, Covelli V, Schiavolin S, Scaratti C, Leonardi M, Willems M. Work-related problems in multiple sclerosis: a literature review on its associates and determinants. Disabil Rehabil 2016 Jul 27;38(10):936-944. [doi: 10.3109/09638288.2015.1070295] [Medline: 26213244]
7. Thomas PW, Thomas S, Kersten P, Jones R, Nock A, Slingsby V, et al. Multi-centre parallel arm randomised controlled trial to assess the effectiveness and cost-effectiveness of a group-based cognitive behavioural approach to managing fatigue in people with multiple sclerosis. BMC Neurol 2010 Jun 16;10(1):43 [FREE Full text] [doi: 10.1186/1471-2377-10-43] [Medline: 20553617]
8. Thomas S, Thomas P, Nock A, Slingsby V, Galvin K, Baker R, et al. Development and preliminary evaluation of a cognitive behavioural approach to fatigue management in people with multiple sclerosis. Patient Educ Couns 2010 Feb;78(2):240-249. [doi: 10.1016/j.pec.2009.07.001] [Medline: 19665337]
9. Thomas S, Thomas PW, Kersten P, Jones R, Green C, Nock A, et al. A pragmatic parallel arm multi-centre randomised controlled trial to assess the effectiveness and cost-effectiveness of a group-based fatigue management programme (FACETS) for people with multiple sclerosis. J Neurol Neurosurg Psychiatry 2013 Oct 21;84(10):1092-1099 [FREE Full text] [doi: 10.1136/jnnp-2012-303816] [Medline: 23695501]
10. Thomas PW, Thomas S, Kersten P, Jones R, Slingsby V, Nock A, et al. One year follow-up of a pragmatic multi-centre randomised controlled trial of a group-based fatigue management programme (FACETS) for people with multiple sclerosis. BMC Neurol 2014 May 19;14(1):109 [FREE Full text] [doi: 10.1186/1471-2377-14-109] [Medline: 24886398]
11. Thomas S, Kersten P, Thomas PW, Slingsby V, Nock A, Jones R, et al. Exploring strategies used following a group-based fatigue management programme for people with multiple sclerosis (FACETS) via the Fatigue Management Strategies Questionnaire (FMSQ). BMJ Open 2015 Oct 20;5(10):e008274 [FREE Full text] [doi: 10.1136/bmjopen-2015-008274] [Medline: 26486976]
12. Thomas S, Pulman A, Thomas P, Collard S, Jiang N, Dogan H, et al. Digitizing a face-to-face group fatigue management program: exploring the views of people with multiple sclerosis and health care professionals via consultation groups and interviews. JMIR Form Res 2019 May 22;3(2):e10951 [FREE Full text] [doi: 10.2196/10951] [Medline: 31120021]
13. Moss-Morris R, McCrone P, Yardley L, van Kessel K, Wills G, Dennison L. A pilot randomised controlled trial of an Internet-based cognitive behavioural therapy self-management programme (MS Invigor8) for multiple sclerosis fatigue. Behav Res Ther 2012 Jun;50(6):415-421. [doi: 10.1016/j.brat.2012.03.001] [Medline: 22516321]
14. Pöttgen J, Moss-Morris R, Wendebourg J, Feddersen L, Lau S, Köpke S, et al. Randomised controlled trial of a self-guided online fatigue intervention in multiple sclerosis. J Neurol Neurosurg Psychiatry 2018 Sep 16;89(9):970-976. [doi: 10.1136/jnnp-2017-317463] [Medline: 29549193]
15. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
16. Online Fatigue Management Course. MS Society. 2019. URL: https://www.mssociety.org.uk/about-ms/signs-and-symptoms/fatigue/managing-fatigue/online-fatigue-management-course [accessed 2021-10-05]
17. Marrie RA, Leung S, Tyry T, Cutter GR, Fox R, Salter A. Use of eHealth and mHealth technology by persons with multiple sclerosis. Mult Scler Relat Disord 2018 Oct 02;27:13-19. [doi: 10.1016/j.msard.2018.09.036] [Medline: 30296732]
18. Van Kessel K, Babbage DR, Reay N, Miner-Williams WM, Kersten P. Mobile technology use by people experiencing multiple sclerosis fatigue: survey methodology. JMIR Mhealth Uhealth 2017 Feb 28;5(2):e6 [FREE Full text] [doi: 10.2196/mhealth.6192] [Medline: 28246073]
19. Griffin N, Kehoe M. A questionnaire study to explore the views of people with multiple sclerosis of using smartphone technology for health care purposes. Disabil Rehabil 2018 Dec;40(12):1434-1442. [doi: 10.1080/09638288.2017.1300332] [Medline: 28322588]
20. Newton A, Bagnell A, Rosychuk R, Duguay J, Wozney L, Huguet A, et al. A mobile phone-based app for use during cognitive behavioral therapy for adolescents with anxiety (MindClimb): user-centered design and usability study. JMIR Mhealth Uhealth 2020 Dec 08;8(12):e18439 [FREE Full text] [doi: 10.2196/18439] [Medline: 33289671]
21. Castle-Clarke S, Curry N, Dorning H, Wetherly L. Improving Care for People With MS: The Potential of Data and Technology. nuffield trust. London: MS Society; 2018. URL: https://www.nuffieldtrust.org.uk/files/2018-07/1531150262_full-web-nuffield-report.pdf
22. Improving Care for People With MS: The Potential of Data and Technology. Summary of Key Conclusions. MS Society. London: MS Society; 2018. URL: https://www.mssociety.org.uk/sites/default/files/2020-08/Improving-care-for-people-with-MS_The-potential-of-data-and-technology_summary.pdf




XSL•FO
RenderX

23. Accelerating Innovation for People Living With MS: Action Plan. London: MS Society; 2018. URL: https://www.mssociety.org.uk/sites/default/files/2020-08/Accelerating-innovation-for-people-living-with-MS-action-plan.pdf
24. Our Research Priorities. MS Society. 2018. URL: https://www.mssociety.org.uk/research/explore-our-research/our-research-priorities [accessed 2021-10-05]
25. Multiple Sclerosis Top 10 Priorities. James Lind Alliance. 2013. URL: http://www.jla.nihr.ac.uk/priority-setting-partnerships/multiple-sclerosis/top-10-priorities/ [accessed 2021-10-05]
26. Giunti G, Guisado-Fernandez E, Caulfield B. Connected health in multiple sclerosis: a mobile applications review. 2017 Presented at: 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS); 2017; Thessaloniki p. 660-665. [doi: 10.1109/cbms.2017.27]
27. Bender JL, Yue RYK, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
28. Plaza I, Demarzo MMP, Herrera-Mercadal P, García-Campayo J. Mindfulness-based mobile applications: literature review and analysis of current features. JMIR Mhealth Uhealth 2013 Nov;1(2):e24 [FREE Full text] [doi: 10.2196/mhealth.2733] [Medline: 25099314]
29. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M, Sainz-De-Abajo B. Comparison of mobile apps for the leading causes of death among different income zones: a review of the literature and app stores. JMIR Mhealth Uhealth 2014 Jan 09;2(1):e1 [FREE Full text] [doi: 10.2196/mhealth.2779] [Medline: 25099695]
30. Babbage DR, van Kessel K, Drown J, Thomas S, Sezier A, Thomas P, et al. MS Energize: field trial of an app for self-management of fatigue for people with multiple sclerosis. Internet Interv 2019 Dec;18:100291 [FREE Full text] [doi: 10.1016/j.invent.2019.100291] [Medline: 31890637]
31. MS Energise. URL: https://msenergize.com [accessed 2020-01-10]
32. van Kessel K, Babbage DR, Kersten P, Drown J, Sezier A, Thomas PW, et al. Design considerations for a multiple sclerosis fatigue mobile app MS Energize: a pragmatic iterative approach using usability testing and resonance checks. Internet Interv 2021 Apr;24:100371 [FREE Full text] [doi: 10.1016/j.invent.2021.100371] [Medline: 33614414]
33. MS Fatigue Fix. URL: https://play.google.com/store/apps/details?id=com.msff&hl=en_US&gl=US [accessed 2021-10-05]
34. Spahrkäs SS, Looijmans A, Sanderman R, Hagedoorn M. Beating cancer-related fatigue with the untire mobile app: protocol for a waiting list randomized controlled trial. JMIR Res Protoc 2020 Feb 14;9(2):e15969 [FREE Full text] [doi: 10.2196/15969] [Medline: 32130185]
35. Untire: beating cancer fatigue together. URL: https://untire.me [accessed 2020-01-10]
36. Spahrkäs S, Looijmans A, Sanderman R, Hagedoorn M. Beating cancer-related fatigue with the Untire mobile app: results from a waiting-list randomized controlled trial. Psychooncology 2020 Nov;29(11):1823-1834 [FREE Full text] [doi: 10.1002/pon.5492] [Medline: 33393199]
37. Giunti G, Guisado FE, Dorronzoro ZE, Rivera RO. Supply and demand in mhealth apps for persons with multiple sclerosis: systematic search in app stores and scoping literature review. JMIR Mhealth Uhealth 2018 May 23;6(5):e10512 [FREE Full text] [doi: 10.2196/10512] [Medline: 29792295]
38. Giunti G, Kool J, Rivera Romero O, Dorronzoro Zubiete E. Exploring the specific needs of persons with multiple sclerosis for mhealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
39. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mhealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
40. Simblett, Evans J, Greer B, Curtis H, Matcham F, Radaelli M, RADAR-CNS consortium. Engaging across dimensions of diversity: a cross-national perspective on mHealth tools for managing relapsing remitting and progressive multiple sclerosis. Mult Scler Relat Disord 2019 Jul;32:123-132 [FREE Full text] [doi: 10.1016/j.msard.2019.04.020] [Medline: 31125754]
41. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
42. Peiris D, Miranda JJ, Mohr DC. Going beyond killer apps: building a better mHealth evidence base. BMJ Glob Health 2018 Feb 21;3(1):e000676. [doi: 10.1136/bmjgh-2017-000676] [Medline: 29527353]
43. Clegg D, Barker R. Case Method Fast-Track: A RAD Approach. Boston: Addison-Wesley Longman Publishing Co; 1994.
44. Fairbanks B, Dogan H, Jiang N, Pretty K, Pulman A, Thomas P, et al. Creating a FACETS digital toolkit to promote quality of life of people with multiple sclerosis through Participatory Design. Proc 32nd Int BCS Human Comput Interact Conf 2018:1-5. [doi: 10.14236/ewic/hci2018.222]
45. Marziniak M, Brichetto G, Feys P, Meyding-Lamadé U, Vernon K, Meuth SG. The use of digital and remote communication technologies as a tool for multiple sclerosis management: narrative review. JMIR Rehabil Assist Technol 2018 Apr 24;5(1):e5 [FREE Full text] [doi: 10.2196/rehab.7805] [Medline: 29691208]
46. Hulme K, Safari R, Thomas S, Mercer T, White C, Van der Linden M, et al. Fatigue interventions in long term, physical health conditions: a scoping review of systematic reviews. PLoS One 2018 Oct 12;13(10):e0203367 [FREE Full text] [doi: 10.1371/journal.pone.0203367] [Medline: 30312325]




XSL•FO
RenderX

47. Beatty L, Binnion C, Kemp E, Koczwara B. A qualitative exploration of barriers and facilitators to adherence to an online self-help intervention for cancer-related distress. Support Care Cancer 2017 Dec;25(8):2539-2548. [doi: 10.1007/s00520-017-3663-2] [Medline: 28299458]
48. Muoio D. Report: Health app supply outstripping demand, leaving small developers behind. Mobi Health News. 2017 Nov 06. URL: https://www.mobihealthnews.com/content/report-health-app-supply-outstripping-demand-leaving-small-developers-behind [accessed 2021-10-05]
49. Material design guidelines. Google. 2020. URL: https://design.google/resources/ [accessed 2021-10-05]
50. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
51. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
52. van der Heijden H. User acceptance of hedonic information systems. MIS Quarterly 2004;28(4):695. [doi: 10.2307/25148660]
53. Android developer guidelines. Google. 2020. URL: https://developer.android.com/guide/index.html [accessed 2021-10-05]
54. Android core app quality guidelines. Google. 2020. URL: https://developer.android.com/docs/quality-guidelines/core-app-quality [accessed 2021-10-05]
55. SAGE Mixed Methods Research. Thousand Oaks: SAGE Publications, Inc; 2019.
56. Brooke J. SUS: a quick and dirty usability scale. Usability Eval Industry 1996;189(194):4-7. [doi: 10.1201/9781498710411-35]
57. Lewis JR. The system usability scale: past, present, and future. Int J Human Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
58. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
59. Sauro J, Lewis J. Quantifying the User Experience: Practical Statistics for User Research. Burlington: Morgan Kaufmann; 2016.
60. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. Journal Of Usability Studies 2009 May 01;4(3):114-123 [FREE Full text]
61. Caelli K, Ray L, Mill J. 'Clear as mud': toward greater clarity in generic qualitative research. Int J Qual Meth 2016 Nov 29;2(2):1-13. [doi: 10.1177/160940690300200201]
62. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
63. Karekla M, Kasinopoulos O, Neto D. Best practices and recommendations for digital interventions to improve engagement and adherence in chronic illness sufferers. Eur Psychol; 24(1)–67 2019;24(1):49-67 [FREE Full text] [doi: 10.1027/1016-9040/a000349]
64. Apolinário-Hagen J, Menzel M, Hennemann S, Salewski C. Acceptance of mobile health apps for disease management among people with multiple sclerosis: web-based survey study. JMIR Form Res 2018 Dec 12;2(2):e11977 [FREE Full text] [doi: 10.2196/11977] [Medline: 30684408]
65. Schwartz B. The Paradox of Choice: Why More Is Less, Revised Edition. New York: Wiley; 2018.
66. Gromisch E, Turner A, Haselkorn J, Lo A, Agresta T. Mobile health (mHealth) usage, barriers, and technological considerations in persons with multiple sclerosis: a literature review. JAMIA Open 2021 Jul;4(3):ooaa067 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa067] [Medline: 34514349]
67. Giunti G, Mylonopoulou V, Rivera Romero O. More Stamina, a gamified mhealth solution for persons with multiple sclerosis: research through design. JMIR Mhealth Uhealth 2018 Mar 02;6(3):e51. [doi: 10.2196/mhealth.9437]
68. Giunti G, Rivera-Romero O, Kool J, Bansi J, Sevillano JL, Granja-Dominguez A, et al. Evaluation of more stamina, a mobile app for fatigue management in persons with multiple sclerosis: protocol for a feasibility, acceptability, and usability study. JMIR Res Protoc 2020 Aug 04;9(8):e18196 [FREE Full text] [doi: 10.2196/18196] [Medline: 32749995]
69. Jongen PJ, ter Veen G, Lemmens W, Donders R, van Noort E, Zeinstra E. The interactive web-based program msmonitor for self-management and multidisciplinary care in persons with multiple sclerosis: quasi-experimental study of short-term effects on patient empowerment. J Med Internet Res 2020 Mar 9;22(3):e14297. [doi: 10.2196/14297]
70. D'hooghe M, Van Gassen G, Kos D, Bouquiaux O, Cambron M, Decoo D, et al. Improving fatigue in multiple sclerosis by smartphone-supported energy management: the MS TeleCoach feasibility study. Mult Scler Relat Disord 2018 May;22:90-96 [FREE Full text] [doi: 10.1016/j.msard.2018.03.020] [Medline: 29649789]
71. Morita P. Design of mobile health technology. Design for Health 2020:87-102. [doi: 10.1016/b978-0-12-816427-3.00005-1]
72. Noorbergen T, Adam M, Roxburgh M, Teubner T. Co-design in mHealth systems development: insights From a systematic literature review. Transactions Human Comput Interactions 2021 Jun 30;13(2):175-205 [FREE Full text] [doi: 10.17705/1thci.00147]
73. Spinuzzi C. The methodology of participatory design. Tech Comm 2005;52(2):163-174 [FREE Full text]
74. Ayobi A, Marshall P, Cox A, Chen Y. Quantifying the body and caring for the mind: self-tracking in multiple sclerosis. Proc 2017 CHI Conf Human Factors Comput Syst 2017 May 02:6889-6901. [doi: 10.1145/3025453.3025869]




XSL•FO
RenderX

75. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You get reminded you're a sick person": personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015 Aug;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
76. Wendrich K, van Oirschot P, Martens MB, Heerings M, Jongen PJ, Krabbenborg L. Toward digital self-monitoring of multiple sclerosis: investigating first experiences, needs, and wishes of people with MS. Int J MS Care 2019 Nov;21(6):282-291 [FREE Full text] [doi: 10.7224/1537-2073.2018-083] [Medline: 31889935]
77. Veenhuizen Y, Cup EH, Jonker MA, Voet NB, van Keulen BJ, Maas DM, et al. Self-management program improves participation in patients with neuromuscular disease. Neurology 2019 Sep 30;93(18):e1720-e1731. [doi: 10.1212/wnl.0000000000008393]
78. Townsend L, Dyer A, Jones K, Dunne J, Mooney A, Gaffney F, et al. Persistent fatigue following SARS-CoV-2 infection is common and independent of severity of initial infection. PLoS One 2020;15(11):e0240784 [FREE Full text] [doi: 10.1371/journal.pone.0240784] [Medline: 33166287]
79. Peters M, Powell V. The Challenge of fatigue. Occup Ther News 2021;29(3):38-39.
80. Pulman A, Taylor J, Galvin K, Masding M. Ideas and enhancements related to mobile applications to support type 1 diabetes. JMIR Mhealth Uhealth 2013;1(2):e12 [FREE Full text] [doi: 10.2196/mhealth.2567] [Medline: 25100684]
81. ShotPut: insulin site rotation. URL: https://apps.apple.com/us/app/shotput-insulin-site-rotation/id916980792 [accessed 2021-10-05]
82. Dexcom Continuous Glucose Monitoring System. URL: https://www.dexcom.com/g6-cgm-system [accessed 2021-10-05]
83. Flash glucose monitoring. Diabetes UK. URL: https://www.diabetes.org.uk/guide-to-diabetes/managing-your-diabetes/testing/flash-glucose-monitoring [accessed 2021-10-05]
84. Alexander S, Peryer G, Gray E, Barkhof F, Chataway J. Wearable technologies to measure clinical outcomes in multiple sclerosis: a scoping review. Mult Scler 2021 Oct 04;27(11):1643-1656 [FREE Full text] [doi: 10.1177/1352458520946005] [Medline: 32749928]
85. White RW, Horvitz E. Cyberchondria. ACM Trans Inf Syst 2009 Nov 01;27(4):1-37. [doi: 10.1145/1629096.1629101]
86. Visvanathan A, Hamilton A, Brady RRW. Smartphone apps in microbiology—is better regulation required? Clin Microbiol Infect 2012 Jul;18(7):E218-E220 [FREE Full text] [doi: 10.1111/j.1469-0691.2012.03892.x] [Medline: 22563840]
87. Hassoon A, Schrack J, Naiman D, Lansey D, Baig Y, Stearns V, et al. Increasing physical activity amongst overweight and obese cancer survivors using an alexa-based intelligent agent for patient coaching: protocol for the Physical Activity by Technology Help (PATH) Trial. JMIR Res Protoc 2018 Feb 12;7(2):e27 [FREE Full text] [doi: 10.2196/resprot.9096] [Medline: 29434016]
88. Corbett T, Singh K, Payne L, Bradbury K, Foster C, Watson E, et al. Understanding acceptability of and engagement with Web-based interventions aiming to improve quality of life in cancer survivors: a synthesis of current research. Psychooncology 2018 Jan;27(1):22-33. [doi: 10.1002/pon.4566] [Medline: 29044832]
89. West R, Michie S. A Guide to Development and Evaluation of Digital Behaviour Interventions in Healthcare. Sutton: Silverback Publishing; 2016.
90. Kang H, Han J, Kwon GH. An ecological approach to smart homes for health care services: conceptual framework of a smart servicescape wheel. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e12425 [FREE Full text] [doi: 10.2196/12425] [Medline: 30735145]
91. Personal Independence Payment. MS Society. URL: https://www.mssociety.org.uk/care-and-support/financial-help/brief-guide-to-benefits/personal-independence-payment/applying-for-pip [accessed 2021-10-05]
92. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016 Dec 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
93. Middleton R, Rodgers W, Chataway J, Schmierer K, Rog D, Galea I, et al. Validating the portal population of the United Kingdom Multiple Sclerosis Register. Mult Scler Relat Disord 2018 Aug;24:3-10. [doi: 10.1016/j.msard.2018.05.015] [Medline: 29860199]
94. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Meth Instr Comput 2003 Aug;35(3):379-383. [doi: 10.3758/BF03195514]

## Abbreviations

**FACETS:** Fatigue: Applying Cognitive Behavioral and Energy Effectiveness Techniques to Lifestyle
**mHealth:** mobile health
**MoSCoW:** must have, should have, could have, and won't have
**MS:** multiple sclerosis
**RCT:** randomized controlled trial
**SMART:** specific, measurable, achievable, and realistic with time for review
**SUS:** System Usability Scale



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 09.04.20; peer-reviewed by G Giunti, L Payne, H James; comments to author 12.06.20; revised version received 29.10.20; accepted 06.09.21; published 09.12.21.*

*Please cite as:*
*Thomas S, Pulman A, Dogan H, Jiang N, Passmore D, Pretty K, Fairbanks B, Davies Smith A, Thomas PW*
*Creating a Digital Toolkit to Reduce Fatigue and Promote Quality of Life in Multiple Sclerosis: Participatory Design and Usability Study*
*JMIR Form Res 2021;5(12):e19230*
*URL: https://formative.jmir.org/2021/12/e19230*
*doi:10.2196/19230*
*PMID:34889744*

©Sarah Thomas, Andy Pulman, Huseyin Dogan, Nan Jiang, David Passmore, Keith Pretty, Beth Fairbanks, Angela Davies Smith, Peter W Thomas. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Mobile Intervention to Improve Sleep and Functional Health of Veterans With Insomnia: Randomized Controlled Trial

Erin Dawna Reilly[1,2,3], MEd, PhD; Stephanie A Robinson[4,5], PhD; Beth Ann Petrakis[4], MPA; Melissa M Gardner[6], MA; Renda Soylemez Wiener[4,5], MPH, MD; Carmen Castaneda-Sceppa[7], MD, PhD; Karen S Quigley[2,4,7], PhD

[1]Mental Illness Research, Education, and Clinical Center, VA Bedford Healthcare System, Bedford, MA, United States
[2]Social and Community Reintegration Research, VA Bedford Healthcare System, Bedford, MA, United States
[3]University of Massachusetts Medical School, Worcester, MA, United States
[4]Center for Healthcare Outcomes and Implementation Research, VA Bedford Healthcare System, Bedford, MA, United States
[5]Pulmonary Division, Boston University School of Medicine, Boston, MA, United States
[6]William James College, Newton, MA, United States
[7]Northeastern University, Boston, MA, United States

**Corresponding Author:**
Erin Dawna Reilly, MEd, PhD
Mental Illness Research, Education, and Clinical Center
VA Bedford Healthcare System
200 Springs Road
Bedford, MA, 01730
United States
Phone: 1 781 687 4191
Email: erin.reilly@va.gov

## Abstract

**Background:** Insomnia is a prevalent and debilitating disorder among veterans. Cognitive behavioral therapy for insomnia (CBTI) can be effective for treating insomnia, although many cannot access this care. Technology-based solutions and lifestyle changes, such as physical activity (PA), offer affordable and accessible self-management alternatives to in-person CBTI.

**Objective:** This study aims to extend and replicate prior pilot work to examine whether the use of a mobile app for CBTI (cognitive behavioral therapy for insomnia coach app [CBT-i Coach]) improves subjective and objective sleep outcomes. This study also aims to investigate whether the use of the CBT-i Coach app with adjunctive PA improves sleep outcomes more than CBT-i Coach alone.

**Methods:** A total of 33 veterans (mean age 37.61 years, SD 9.35 years) reporting chronic insomnia were randomized to use either the CBT-i Coach app alone or the CBT-i Coach app with a PA intervention over 6 weeks, with outcome measures of objective and subjective sleep at pre- and posttreatment.

**Results:** Although the PA manipulation was unsuccessful, both groups of veterans using the CBT-i Coach app showed significant improvement from baseline to postintervention on insomnia ($P<.001$), sleep quality ($P<.001$), and functional sleep outcomes ($P=.002$). Improvements in subjective sleep outcomes were similar in those with and without posttraumatic stress disorder and mild-to-moderate sleep apnea. We also observed a significant but modest increase in objective sleep efficiency ($P=.02$).

**Conclusions:** These findings suggest that the use of a mobile app–delivered CBTI is feasible and beneficial for improving sleep outcomes in veterans with insomnia, including those with comorbid conditions such as posttraumatic stress disorder or mild-to-moderate sleep apnea.

**Trial Registration:** ClinicalTrials.gov NCT03305354; https://clinicaltrials.gov/ct2/show/NCT03305354

***(JMIR Form Res 2021;5(12):e29573)*** doi:10.2196/29573

**KEYWORDS**

cognitive behavioral therapy; mobile app; physical activity; insomnia



XSL•FO
RenderX

## Introduction

### Background

Sleep disturbances, especially chronic insomnia (difficulty falling and staying asleep), are among the most prevalent complaints following military deployments, particularly in post-9/11 veterans of Operations Enduring Freedom, Iraqi Freedom, and New Dawn [1-4]. According to an evaluation of medical records of approximately 10 million US veterans, 91% were prescribed at least one sleep medication, with sleep apnea (47%) and insomnia (26%) being two of the most common diagnoses [5]. There are both service-related and psychological reasons why sleep disturbances can emerge during deployment and endure past active military service. These include unusual sleep–wake schedules because of military work schedules, separation from family and loved ones, poor sleeping conditions that are detrimental to good sleep hygiene, and threat of or actual injury to self or others and associated subjective arousal [6,7]. Moreover, military service can have additional repercussions for postdeployment reintegration into civilian life [8], including greater risk for specific psychiatric comorbidities of sleep disturbance such as posttraumatic stress disorder (PTSD) [9], depression [10], and chronic pain [11].

The co-occurrence of mental and physical symptoms with sleep disturbance can add complexity to this health issue. For instance, evidence now suggests that sleep disturbance can precede or follow the onset of depression, suggesting that in some cases, sleep disturbance can increase the likelihood of depressive symptoms [12-14], including in veterans [10]. It is also possible that both sleep disturbance and depression are different manifestations of an underlying metabolic or energy regulatory dysfunction, as it has been theorized that energy regulation is the key function of all nervous systems [15,16].

Regardless of the root cause, sleep disturbance in post-9/11 veterans also commonly co-occurs with poor functional health and reduced social and community engagement [17-19]. Chronic insomnia is a transdiagnostic health problem affecting physical, cognitive, and emotional functioning [4,20], which in turn can complicate veteran community reintegration (CR) or the ability to return to family, vocational, and community life [21]. Research on insomnia suggests that poor sleep can negatively affect specific areas of CR, such as academic functioning issues [22], employment success [23], and social relationships related to loneliness and feelings of belonging [24]. Poor functioning and CR are a particularly serious and common problem for veterans with mental health concerns who frequently do not themselves seek out or avail health care [25,26]. This further affects the difficulty of treating poor sleep, as the confluence of insomnia and comorbid mental health concerns in veterans together can negatively affect treatment efficacy for sleep outcomes [27,28].

### Cognitive Behavioral Therapy for Insomnia

The use of over-the-counter sleep aids and prescription medications is common and remains a frontline treatment for insomnia [29]. However, meta-analytic research is mixed, with some showing that pharmacological and nonpharmacological insomnia interventions have comparable efficacy [30] and others suggesting that behavioral interventions show greater improvements in sleep quality than pharmacological interventions [31]. The effects of nonpharmacological interventions tend to be more durable than sleep medications, with treatment gains persisting after treatment [32].

Consequently, cognitive behavioral therapy for insomnia (CBTI) is considered the gold standard nonpharmacological treatment for insomnia in terms of both effectiveness and efficacy [32-34]. This manualized treatment implements multiple cognitive and behavioral techniques, including sleep restriction, psychoeducation, and cognitive restructuring, and has been evaluated as a frontline intervention among veterans [35,36]. However, for veterans facing competing demands from work, school, family, and other health concerns, adherence to in-clinic CBTI treatment can be poor [37]. A major barrier to in-person CBTI treatment is its time-intensive nature, which reduces both accessibility and adherence [7].

Mobile apps delivering self-management–based, personalized CBTI have the potential to reduce these barriers and assist with in-person treatments [38-40]. To capitalize on this delivery option, the cognitive behavioral therapy for insomnia coach app (CBT-i Coach app; US Department of Veterans Affairs) was launched in 2013 [41]. Although the app was not intended to completely replace CBTI provider–delivered interventions, it is a valuable adjunct to provide sleep-management tools and educational resources and, in particular, to track self-reported sleep metrics and subjective sleep. Although existing research is preliminary, the CBT-i Coach app has shown promise in positively affecting sleep when used in coordination with individual CBTI treatments [42] and as a stand-alone self-management option [43].

In addition to preliminary, positive research findings on self-management of insomnia via mobile sleep apps, interest in physical activity (PA) for insomnia treatment has also increased as clinicians look for conjunctive behavioral interventions for sleep improvement. Often, treatment as usual for insomnia includes a health care provider's recommendation for improving lifestyle choices such as eating a healthier diet and increasing exercise as a way to promote better sleep [7]. Such routine clinician suggestions for increasing PA are based on research studies suggesting that taking more daily steps is positively associated with both longer sleep duration and better sleep quality at night [44]. Even modest increases in PA have been proposed to provide an easily accessible, nonpharmacological treatment alternative for sleep disturbance [45]. Specifically, some research suggests that long-term use of moderate aerobic exercise can improve sleep quality and functioning in individuals with insomnia [46] and obstructive sleep apnea (OSA) [47]. However, a recent meta-analysis found both positive and negative associations between daily PA and sleep duration [48], suggesting that other factors such as insomnia severity and mental health comorbidities may affect the direction of an often-assumed positive relationship between exercise and sleep. Given the common recommendation for increased PA in patients with insomnia, these findings suggest that future research is needed to actively explore what impact, if any, PA can have for patients already using a common insomnia self-management program.



XSL•FO
RenderX

### This Study

In a recent pilot project, we used a combination of the CBT-i Coach app and a home-based objective sleep monitor to assess and provide feedback to post-9/11 veterans regarding their home-based sleep behaviors [40]. The CBT-i Coach app had high usability, and veterans with insomnia showed improvements in both self-reported and sleep monitor–assessed sleep quality. This study also revealed that a large proportion of these veterans (54%) unexpectedly screened positive for OSA using the WatchPAT (Itamar Medical, Inc) at-home sleep monitor, a sleep monitor validated for detecting OSA [49]. As we excluded individuals for whom either of the 2 initial home-based sleep tests suggested greater than mild sleep apnea, this pilot was unfortunately unable to characterize the impact of the mobile sleep intervention on their insomnia. In addition, prior research has suggested that a PA adjunct could potentially provide additional benefits to sleep outcomes in veterans, even for those with mild-to-moderate sleep apnea.

This project seeks to replicate initial pilot results on sleep outcomes, while extending prior work to include veterans with mild-to-moderate sleep apnea. We also assess the outcomes of PA as an adjunctive treatment in a second intervention group that used the CBT-i Coach app and guide them to increase their PA via increased daily steps (PA). Thus, we conduct a small single-blind, pilot randomized controlled trial (RCT; randomized 1:1) with the following 2 arms: (1) CBT-i Coach app use only (CBT-i only) and (2) CBT-i Coach app use plus a PA intervention (CBT-i +PA). We examine the following a priori, preregistered Clinical Trials.gov (identifier: NCT03305354) for sleep hypotheses and analyses:

1. The group using the CBT-i Coach app for 6 weeks, with an additional PA intervention component (CBT-i +PA), would report better subjective sleep outcomes (self-reported insomnia severity, sleep quality, and functional sleep) and objective sleep outcomes (sleep efficiency) compared with participants using only the CBT-i Coach app.
2. There would be pre- to postintervention improvements in subjective sleep outcomes, objective sleep efficiency, and functional measures for all participants (both CBT-i +PA and CBT-i only).

In addition, we examine 2 a priori registered hypotheses related to functional health and community engagement:

1. There would be improvements in self-reported community integration and functioning outcomes over the course of the trial for each group and for the full sample.
2. There would be improvements in subjective sleep as reported by participants in the CBT-i Coach app sleep diaries for total sleep, sleep efficiency, and self-reported sleep quality, for each group and for the full sample.

## Methods

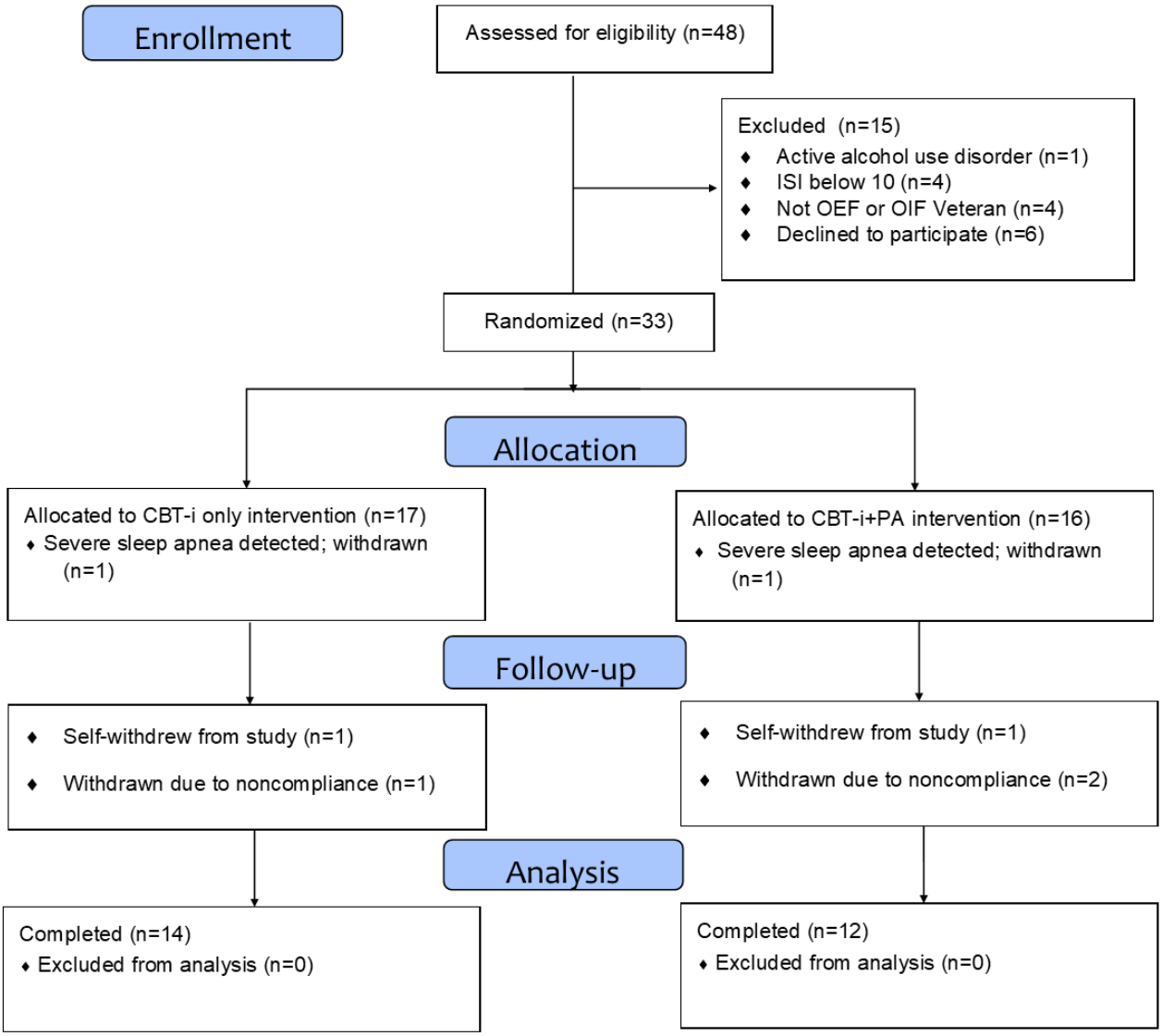
### Participant Recruitment

The study was approved by the institutional review board of the VA Bedford Healthcare System, Bedford, Massachusetts. Veterans were recruited via flyers, presentations, community outreach, provider referrals, and recruitment letters. Interested veterans were phone screened for study eligibility. The participant flow has been described in Figure 1.



XSL•FO
RenderX

**Figure 1.** CONSORT diagram of participant inclusion and attrition. CBT-i: cognitive behavioral therapy for insomnia; ISI: insomnia severity index; OEF: Operation Enduring Freedom; OIF: Operation Iraqi Freedom; PA: physical activity.




## Participant Demographics

The mean age of the randomized sample was 37.61 years (SD 9.35 years), and the mean BMI was 28.03 (SD 4.27; ie, overweight). Participants (25/33, 76% men and 8/33, 24% women) had an average baseline insomnia severity index (ISI) of 17.91 (SD 4.05), indicating moderately severe clinical insomnia. Approximately 76% (25/33) identified as White and 18% (6/33) identified ethnically as Hispanic. No participants had been diagnosed with apnea before enrollment; however, the WatchPAT sleep assessment (completed by 32/33, 97% of the enrolled participants) revealed that 38% (12/32) had an apnea–hypopnea index (AHI) score corresponding to no or minimal apnea, 25% (8/32) had mild apnea, 31% (10/32) had moderate apnea, and 6% (2/32) had severe apnea. In addition, none of the participants had previously used the CBT-i Coach app (0/33, 0% of participants), and only some (4/33, 12%) participants had previously used the Fitbit (Fitbit Inc) tracker and mobile app (see Table 1 for detailed demographic information).



XSL•FO
RenderX

**Table 1.** Descriptive statistics at baseline for participants by group (N=33).

| Characteristics | CBT-i Coach only group (n=17) | CBT-i Coach+physical activity (n=16) | Total |
|---|---|---|---|
| Age (years), mean (SD) | 36.24 (8.74) | 39.06 (10.00) | 37.61 (9.35) |
| BMI (kg/m$^2$), mean (SD) | 26.84 (4.43) | 29.30 (4.12) | 28.03 (4.27) |
| Baseline ISI[a], mean (SD) | 18.00 (3.57) | 17.81 (4.57) | 17.91 (4.05) |
| Baseline weekly steps, mean (SD) | 9487 (5628) | 9942 (4841) | 9697.14 (5181) |
| **Gender, n (%)** | | | |
| Male | 12 (71) | 13 (81) | 25 (76) |
| Female | 5 (29) | 3 (19) | 8 (24) |
| **Apnea rates[b], n (%)** | | | |
| None or minimal | 5 (31) | 7 (44) | 12 (38) |
| Mild | 4 (25) | 4 (25) | 8 (25) |
| Moderate | 6 (38) | 4 (25) | 10 (31) |
| Severe | 1 (6) | 1 (6) | 2 (6) |
| **Race[b], n (%)** | | | |
| White/Caucasian | 13 (76) | 12 (75) | 25 (76) |
| Black/African American | 2 (12) | 3 (19) | 5 (15) |
| Japanese | 1 (6) | 0 (0) | 1 (3) |
| Middle Eastern | 1 (6) | 0 (0) | 1 (3) |
| American Indian | 1 (6) | 0 (0) | 1 (3) |
| Other | 2 (12) | 1 (6) | 3 (9) |
| **Ethnicity, n (%)** | | | |
| Hispanic/Latino | 4 (24) | 2 (13) | 6 (18) |
| Not Hispanic/Latino | 13 (76) | 14 (87) | 27 (82) |
| **Income (US $)[b], n (%)** | | | |
| <11,999 | 2 (12) | 1 (7) | 3 (9) |
| 12,000-24,999 | 3 (18) | 0 (0) | 3 (9) |
| 25,000-49,999 | 7 (41) | 4 (27) | 11 (34) |
| 50,000-99,000 | 1 (6) | 3 (20) | 4 (13) |
| ≥100,000 | 4 (24) | 7 (47) | 11 (34) |

[a]ISI: insomnia severity index.

[b]Not required to answer or missing data point; numbers may not add up to the total sample of N=33.

## Description of the Interventions: CBT-i Coach App, Self-management Guide, Fitbit, and WatchPAT

### *Overview*

For 6 weeks, all participants were asked to use the CBT-i Coach app, to which we added self-management support. Participants were randomized to either (1) use of the CBT-i Coach app alone (CBT-i alone) or (2) CBT-i Coach plus PA intervention (CBT-i+PA). Those in the CBT-i +PA group were asked to use the CBT-i Coach app and increase their average daily step count by 10% each week or until they walked 10,000 steps a day. All participants were loaned a Fitbit Charge 2, WatchPAT, and iPod Touch on which the CBT-i Coach app was installed (iOS version 2.3). Participants wore the WatchPAT 3 times: (1) one night soon after study enrollment; (2) a second time, generally within the first week after enrollment; and (3) one night after they had completed the 6-week intervention. On average, the time between participants' first and last WatchPAT data collection night was 44 days.

### *CBT-i Coach App*

The CBT-i Coach app offers sleep psychoeducation, tools for tracking sleep (eg, daily sleep diaries), and sleep hygiene recommendations, including cultivating a conducive sleep environment, engaging in regular exercise, and maintaining a healthy diet. In-app relaxation tools include guided imagery audio clips, tips for winding down, breathing tools, and progressive muscle relaxation. A behavioral plan can also be reviewed and updated in the CBT-i Coach app, including setting



XSL•FO
RenderX

reminders for when to go to sleep and get out of bed, completing sleep diaries, taking insomnia assessments, engaging in scheduled worry time, and stopping caffeine intake for the day. The CBT-i Coach app allows users to see graphical depictions of their sleep diary data and ISI scores. Additional CBTI worksheets were embedded in a separate in-house–created app (Sleep Help app) and included wakeful activities, coping self-statements, constructive worry, and a relaxation log similar to worksheets from the workbook *Quiet Your Mind and Get To Sleep* [50]. The daily input of a sleep diary required approximately 2 to 3 minutes per day to complete. The amount of time that individual participants allotted to additional CBT-i Coach app activities that were suggested (but not required) for participants via the self-management guide (see the following section) varied. We estimated that these required an additional 3 (eg, completing a sleep assessment) to 6 minutes (eg, progressive muscle relaxation meditation).

### *Self-management Guide*

A self-management guide (Multimedia Appendix 1; CBT-i +PA group version), available on both the iPod Touch and paper copy, provided week-by-week suggestions for using elements of the app and the worksheets. For each week of the 6-week intervention, the guide suggested what materials to read in the app; which features to use, such as completing a daily sleep diary each morning; and which worksheets to complete. Participants in the CBT-i +PA arm were provided with additional instructions, guidelines, and resources to help motivate and support their weekly step increases. Specifically, their version of the self-management guide reminded these participants to increase their daily step counts every week, reminded them to track their steps daily in their paper log, and requested that they *boost* their step counts by 10% weekly. They were also provided with a 24-page researcher-modified motivational and instructional guide titled *Stepping Out*, which was based on prior research, clinical guidelines, and national walking guidebooks (ie, *Stepping Out Mature Adults: Be Healthy, Walk Safely* [51]). Completion of these PA-specific tasks added approximately 8 to 10 minutes per week to their activities. This PA intervention was purposefully *light touch* and involved gentle motivational reminders, simulating common real-world clinical suggestions given to patients and allowed us to assess the feasibility of a PA adjunct for a possible future implementation trial.

### *WatchPAT Sleep Monitor*

Objective sleep was recorded using a WatchPAT (model WP200U) sleep monitor (Itamar Medical, Inc). The WatchPAT sleep monitor is a Food and Drug Administration–approved device that assesses objective sleep parameters and is valid for screening measures of OSA [49] and broad sleep stage measures (eg, rapid eye movement [REM] vs non-REM) as compared with polysomnography-based measures [52]. The WatchPAT is worn like a wristwatch attached via a cable to a plethysmographic-based finger-mounted probe and a small sensor on the chest to measure snoring. It is less obtrusive and less disruptive of sleep than in-laboratory or in-home polysomnography. Participants can use the device with simple instructions, which were provided via a video on the iPod Touch and a laminated pamphlet. Participants also self-recorded any prescription medications and nonprescription medications, including vitamins taken during the 24 hours before each WatchPAT night.

### *PA Monitor*

PA was recorded with a Fitbit Charge 2, a wireless-enabled tool that provides measures of heart rate, quality of sleep, number of steps walked, steps climbed, and estimates calories burned. It is worn like a wristwatch and has been well-validated for use in measuring steps [53].

## Measures

### *Overview*

Participants had to have served in the military since 2001 in Iraq or Afghanistan and report current insomnia defined by an ISI score>10 with a duration of at least 1 month and impaired daytime functioning as measured by endorsing *much* or *very much* on ISI item 7. This ISI cutoff was chosen from prior research, which found that a total score of ≥11 on the ISI indicated insomnia disorder in clinical samples with 97.2% sensitivity and 100% specificity [54]. As some research has suggested that insomnia is underdiagnosed in the veteran population [5], participants were not required to have a formal, prior diagnosis of insomnia disorder within their electronic medical records. Participants were excluded if they demonstrated moderate-to-severe cognitive impairment on the Telephone Mini Mental State Exam [55] and excessive alcohol use on the Alcohol Use Disorders Identification Test-Concise [56] or reported periodic leg movement symptoms or a circadian rhythm disorder. As subjective and objective measures of sleep often do not correlate highly in patients with insomnia [57] and can differentially predict treatment outcomes [54], we followed best practices and measured the impact of our intervention on both subjective and objective sleep outcomes.

### *Subjective Sleep Measurement*

Self-reports of insomnia, sleep quality, and functional outcomes because of sleep were measured at baseline and final assessment visits using the ISI, the Pittsburgh Sleep Quality Index (PSQI), and the Functional Outcomes of Sleep Questionnaire–10 items (FOSQ-10), respectively. The ISI has been shown to be sensitive to changes in insomnia severity with CBTI interventions [58,59], with possible scores ranging from 0 to 28 and higher scores indicating more severe insomnia. The PSQI, a global measure of perceived sleep quality, has been extensively used in sleep trials and has been shown to be sensitive to changes after CBTI [37,59]. Scores on the PSQI range from 0 to 21, with a higher score indicating worse sleep quality. The FOSQ-10 [60] was used to assess the impact of sleepiness on functioning in everyday activities. Possible scores range from 5 to 20, with higher scores indicating a better functional status. In this sample, Cronbach α were acceptable (ISI, Cronbach α=.70; PSQI, Cronbach α=.75; and FOSQ-10, Cronbach α=.82). Sleep diary data were collected through the CBT-i Coach app and included self-reported total sleep time, sleep quality (1=very poor to 5=very good), and app-calculated sleep efficiency (percentage of time spent asleep while in bed).



XSL•FO
RenderX

### *Objective Sleep Measurement*

Objectively measured sleep efficiency was obtained from the WatchPAT across 3 nights. To control for first night effects, we used the second WatchPAT sleep assessment night as the preintervention sleep efficiency (on average, approximately 8 days after enrollment). The WatchPAT calculates the proportion of REM sleep using a genetic algorithm (ie, a machine learning technique) to determine REM sleep onset and offset. The WatchPAT recorded total sleep time, total and percent time spent in light, deep and REM sleep stages, AHI, respiratory disturbance index, and number of awakenings; for this study, sleep efficacy was the primary measure of interest.

### *Functioning and Mental Health Outcomes*

Additional outcomes included social engagement, community engagement, and health-related quality of life. The Lubben Social Network Scale–6 items (LSNS-6) [61] is a 6-item scale measuring 5 aspects of social isolation and social networks, with scores ranging from 0 to 30 and higher scores indicating lower levels of social isolation. The Community Integration Questionnaire (CIQ) [62] is a 15-item self-report measure that assesses productivity and engagement in activities in the home and in social settings, with a range of 0 to 39 points and greater scores indicating greater community engagement. Finally, the Veterans RAND 12 Item Health Survey (VR-12) [63] is a reliable and effective measure that documents perceived change in veterans' health-related quality of life. It provides both a mental component score (MCS) and a physical component score (PCS) ranging from 0 to 100, with higher scores indicating better functioning. The LSNS-6 ($\alpha$=.87) and VR-12 ($\alpha$=.94) showed good internal reliability. The CIQ showed poor internal reliability in this sample ($\alpha$=.40), and as such, CIQ results should be interpreted with caution.

Additional measures were used to assess whether our randomization process resulted in balanced groups in terms of mental health and physical symptoms. These were evaluated using the Patient Health Questionnaire (PHQ)-15 [64], depressive symptoms using the PHQ-9 [65] and pain disability index [66], and PTSD symptoms with the PTSD checklist (PCL-5) [67]. Each of these measures demonstrated good internal consistency (PHQ-15: $\alpha$=.83; PHQ-9: $\alpha$=.82; pain disability index: $\alpha$=.82; and PCL-5: $\alpha$=.95).

## Procedure

After a brief phone screen to assess for potential eligibility based on veteran status, age, and insomnia severity, participants were scheduled for an initial study visit to complete the consent process, eligibility testing, and baseline study activities. At the initial study visit, participants were provided with an informed consent document detailing the study expectations, procedures, requirements for participation (including use of different technologies such as the CBT-i Coach app, Fitbit, and WatchPAT), assessment schedule, and compensation. The informed consent document and Health Insurance Portability and Accountability Act authorization form were also reviewed verbally by a research staff member (BAP, EDR, or SAR) before being signed by the participant. The potential participants were then provided with an opportunity to ask any questions. Only after all questions were answered did the participants sign the written consent to participate and complete self-report measures and surveys. Participants were then randomly assigned to CBT-i alone or CBT-i +PA intervention. A researcher demonstrated the use of the iPod Touch, CBT-i Coach app, Fitbit, WatchPAT, and Sleep Help app. All participants were instructed to complete daily sleep diaries, complete activities in the self-management guide, and wear the WatchPAT once during the next week. A follow-up visit was scheduled for the participant to return after wearing the WatchPAT to assess whether participants screened positive for severe apnea. Individuals randomized to the CBT-i +PA condition were given a step-tracking document, an adapted handbook for safely increasing one's steps [68], and instructions to increase their step count by 10% each week over the 6 weeks.

During the second visit, a researcher downloaded and printed the data from the WatchPAT sleep report. Individuals with no apnea diagnosis or a positive screen for mild or moderate OSA (AHI<30/hour of sleep) at either the first or second WatchPAT testing were retained in the trial, and those with severe OSA (AHI>30/hour of sleep) were excluded from further participation and referred to a primary care provider. Given the well-known *first night effect* in which sleep can be negatively affected by sleep monitoring [69], particularly in those with insomnia [70], participants then completed a second night of WatchPAT monitoring within the coming week, which was reviewed again. At the third and final in-person visit (ie, postintervention), participants returned all devices, completed a postintervention survey battery and interview, and received their final WatchPAT report. Participants received US $15 for the 2 in-person assessment screenings during preintervention, US $15 for a midintervention phone assessment, and US $400 for the postintervention assessment.

## Power Analysis

To meet our original pilot goal of ultimately calculating a reliable effect size for a later fully-powered RCT, we conducted a priori power analyses for the proposed nonparametric analyses (using GPower 3.1 [71]). For an effect size Cohen f=0.3 (moderate via Cohen conventions; $\alpha$=.05, 1-$\beta$=.80, 2 groups with 5 potential covariates), we would need N=45 for each of the 2 groups. For this pilot RCT, we proposed that N=15 per group would provide a reasonable estimate of the effect sizes for each group for the future RCT, as numerous such studies appear in the peer-reviewed literature with similar group numbers (eg, group N=10, 11, or 17 [72])

## Data Analysis Plan

We first assessed the potential differences between groups on sleep, functioning, and mental health outcomes. In addition, we completed a planned manipulation check to ensure that participants in the CBTI+PA group increased their steps relative to the CBTI alone group. To address our a priori hypotheses, we evaluated whether subjective (ISI, PSQI, and FOSQ-10) and objective sleep efficiency changed from preintervention to postintervention using analyses suitable for our small sample size, namely Wilcoxon signed rank tests and Mann–Whitney *U* tests comparing the medians of distributions. To address our functioning and mental health hypotheses, we used Wilcoxon signed rank tests to assess changes between baseline and



XSL•FO
RenderX

postintervention on the LSNS-6, CIQ, and VR-12 MCS, and PCS scores. Outcomes were analyzed using an intent-to-treat approach, which included all randomized participants. To account for missing data, we used a *last observation called forward* method so we could use the last completed follow-up as outcome data for all participants not removed for meeting exclusion criteria (eg, severe apnea) or initial protocol noncompliance (eg, not completing the full baseline procedures for intervention onboarding).

# Results

## Preliminary Analyses

Nonparametric Wilcoxon signed rank tests revealed no significant group differences between the CBTI alone and CBT-i+PA groups on baseline subjective and objective sleep measures (ISI, PSQI, FOSQ-10, and sleep efficiency). The daily use of the CBT-i Coach app was high; participants on average completed sleep diaries for 86% (36/42 days) of study nights. In addition, there were no significant between-group differences in baseline mental and physical health on the VR-12. However, importantly, groups differed significantly at baseline on the PCL-5 ($P$=.03, with a median score of 20.50 for the CBTI alone group and a median score of 42.50 for the CBTi+PA group. Thus, post hoc analyses were conducted to assess the potential impact of the differences in PTSD symptoms on sleep outcomes.

Manipulation check analyses revealed that there were no significant differences between baseline steps (mean 10,193, SD 6161) and postintervention steps (mean 8382, SD 4629) for the CBTI alone group (Z=−0.80; $P$=.41) and no significant differences between baseline steps (mean 9485, SD 4798) and postintervention steps (mean 7653, SD 3941) for the CBT-i+PA group (Z=−1.96; $P$=.05). In addition, the CBT-i+PA group did not significantly increase their steps compared with the CBTI alone group (Z=−0.92; $P$=.37), indicating a failure of the PA manipulation.

## Sleep Outcomes

There were no significant differences between the CBT-i alone group and the CBT-i+PA group from preintervention to postintervention for the ISI ($U$=66; $P$=.37), PSQI ($U$=76; $P$=.98), FOSQ-10 ($U$=77.5; $P$=.74), or sleep efficiency ($U$=54; $P$=.60; see Table 2 for mean, median, and SD descriptive information by group).

Across all participants using the CBT-i Coach app (ie, without considering group and per a priori hypothesis 2), there was a significant decrease in ISI scores (Z=−4.31; $P$<.001) from preintervention (median 16.00) to postintervention (median 11.00), which reflects a decrease from moderate clinical insomnia (ISI range 15-21) to subthreshold insomnia (ISI range 8-14), although the 5-point median decrease was below the suggested 6-point cutoff for clinical improvement. Of the 26 final participants, 24 (92%) reported reduced insomnia symptoms over the 6-week intervention, and 2 (8%) reported increased insomnia symptoms. There was also a significant decrease in PSQI scores (Z=−3.57; $P$<.001) from preintervention (median 14.00) to postintervention (median 9.00), with 77% (20/26) of participants reporting improved sleep quality, 15% (4/26) reporting their sleep as worse, and 8% (2/26) reporting no difference. The median PSQI score at postintervention (median 9.00) fell below the suggested cutoff for clinically significant insomnia for veterans (cutoff=10).

There was also a significant increase in FOSQ-10 scores (Z=3.13; $P$=.002) from preintervention (median 14.00) to postintervention (median 15.58), with 77% (20/26) of participants reporting better functional sleep outcomes and 23% (6/26) reporting worse functioning. The median improvement of 1.58 points was slightly below the suggested clinically significant minimal important difference (1.7-2.0), although this minimal important difference is suggested for participants with greater insomnia severity because of narcolepsy or diagnosed OSA. Given the high rate of sleep apnea, we also ran a post hoc analysis to investigate the impact of baseline apnea on changes in sleep outcomes. Those without apnea reported significant improvement in their ISI (Z=−2.807; $P$=.005), as did participants with mild to moderate sleep apnea ($Z$=−3.19; $P$<.001), suggesting that mild to moderate apnea did not prevent intervention-related ISI improvement.

Using available WatchPAT data (n=23), we observed a significant increase in sleep efficiency ($Z$=−2.4; $P$=.02) from preintervention (median 84.13) to postintervention (median 85.32). Specifically, 65% (15/23) of participants showed an increase in sleep efficiency over the 6-week intervention from the second WatchPAT monitor assessment to the last, whereas 35% (8/23) showed a reduction in overall sleep efficiency. In addition, post hoc Friedman tests for nonparametric repeated measures were used to compare total sleep times, sleep efficiency, and sleep quality reported in the CBT-i Coach app sleep diary across the 6 weeks of the trial for the full sample (CBT-i and CBT-i +PA). Using this self-reported data, there were no significant differences across weekly averages for total sleep ($\chi^2_{5,22}$=2.3; $P$=.81) or sleep efficiency ($\chi^2_{5,22}$=9.3; $P$=.10). However, there was a significant increase in reported sleep quality across the 6 weeks ($\chi^2_{5,22}$=12.6; $P$=.03), with post hoc Bonferroni-adjusted tests revealing a significant increase in sleep quality from week 1 (median 2.64) to week 5 (median 3.14; $P$=.03).



XSL•FO
RenderX

**Table 2.** Outcome measures at preintervention and postintervention.

| Measures | Scale range | CBT-i[a] only group (n=14) | | | | *P* value | CBT-i+physical activity group (n=12) | | | | *P* value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Preintervention | | Postintervention | | | Preintervention | | Postintervention | | |
| | | Values, mean (SD) | Values, median (range) | Values, mean (SD) | Values, median (range) | | Values, mean (SD) | Values, median (range) | Values, mean (SD) | Values, median (range) | |
| ISI[b] | 0-28 | 17.64 (3.77) | 16.00 (13.00-26.00) | 10.86 (4.24) | 11.00 (3.00-22.00) | <.001 | 17.92 (4.94) | 17.00 (12.00-26.00) | 13.08 (5.09) | 13.50 (4.00-23.00) | .004 |
| PSQI[c] | 0-21 | 13.07 (3.29) | 13.50 (8.00-19.00) | 9.86 (4.24) | 9.50 (3.00-18.00) | .004 | 12.63 (4.52) | 14.00 (4.00-18.00) | 10.50 (4.56) | 9.50 (2.00-19.00) | .03 |
| FOSQ-10[d] | 5-20 | 13.67 (3.50) | 14.33 (5.83-19.00) | 15.57 (3.27) | 16.25 (9.50-19.33) | .04 | 12.67 (3.22) | 12.00 (7.67-17.17) | 14.90 (2.53) | 15.44 (10.33-18.50) | .02 |
| LSNS-6[e] | 0-30 | 11.64 (7.40) | 11.50 (1.00-21.00) | 11.71 (7.13) | 12.00 (2.00-22.00) | .84 | 9.42 (4.36) | 9.00 (1.00-19.00) | 10.00 (4.45) | 10.00 (3.00-19.00) | .52 |
| CIQ[f] | 0-39 | 28.43 (3.00) | 29.00 (22.00-32.00) | 27.57 (3.98) | 28.00 (21.00-33.00) | .19 | 27.42 (3.42) | 28.00 (21.00-32.00) | 29.00 (3.57) | 29.50 (23.00-34.00) | .15 |
| VR-12[g] PCS[h] | 0-100 | 44.70 (9.04) | 43.74 (22.34-56.61) | 43.44 (9.06) | 43.43 (24.31-56.85) | .30 | 41.81 (11.32) | 41.06 (23.67-57.19) | 40.85 (8.59) | 40.85 (26.36-54.46) | .75 |
| VR-12 MCS[i] | 0-100 | 38.04 (13.73) | 38.84 (18.87-56.61) | 44.05 (13.63) | 40.26 (26.54-64.35) | .02 | 39.97 (10.93) | 39.11 (19.48-58.32) | 38.49 (14.60) | 39.75 (14.37-62.97) | .53 |

[a]CBT-i: cognitive behavioral therapy for insomnia coach app.

[b]ISI: insomnia severity index.

[c]PSQI: Pittsburgh Sleep Quality Index.

[d]FOSQ-10: Functional Outcomes of Sleep Questionnaire–10.

[e]LSNS-6: Lubben Social Network Scale–6 items.

[f]CIQ: Community Integration Questionnaire.

[g]VR-12: Veterans RAND 12 Item Health Survey.

[h]PCS: physical component score.

[i]MCS: mental component score.

## Functioning and Mental Health Outcomes

Mann–Whitney *U* tests revealed no significant differences between groups in changes from preintervention to postintervention for the LSNS-6 (*U*=75; *P*=.67), the VR-12 PCS (*U*=78; *P*=.78), or CIQ (*U*=46; *P*=.05). However, there was a significant between-group difference in the VR-12 MCS (*U*=44; *P*=.04). Specifically, the CBT-i +PA group had a statistically significant but numerically small and clinically marginal decrease in VR-12 MCS scores from preintervention (median 40.2) to postintervention (median 39.7; see Table 2 for descriptive information by group). Across groups, there were no significant improvements in the LSNS-6 (*P*=.54), CIQ (*P*=.88), VR-12-PCS (*P*=.15), or VR-12 MCS (*P*=.17).

## Additional Post Hoc Analyses of PTSD and Sleep Outcomes

As preliminary analyses revealed a failure of randomization such that by random chance, there was significantly higher PTSD symptomology in the CBT-i + PA group, we conducted post hoc analyses to evaluate the potential impact of PTSD symptom severity on primary sleep outcomes. Compared with participants below the symptom cutoff suggesting no clinically relevant PTSD symptomology, individuals with PTSD symptom severity in the clinical range (as indicated by a score>33 on the PCL-5) had significantly worse subjective sleep pre intervention (Table 3), as reflected in their baseline ISI (*Z*=−1.99; *P*=.04), PSQI (*Z*=−2.18; *P*=.03), and FOSQ-10 (*Z*=−2.74; *P*=.01). However, those with and without clinically significant PTSD symptoms showed significant improvements from preintervention to postintervention on self-reported sleep outcomes. Specifically, veterans with PTSD symptoms in the clinical range reported significant improvements in their ISI (*Z*=−2.76; *P*=.01), FOSQ-10 (*Z*=−2.40; *P*=.02), and PSQI (*Z*=−2.19; *P*=.03) scores, as did the group without clinically significant PTSD symptom severity on the ISI (*Z*=−3.21; *P*<.001), FOSQ-10 (*Z*=−2.29; *P*=.02), and PSQI (*Z*=−2.92; *P*=.01). Using the PCL-5 clinical cutoff scores, we also



XSL•FO
RenderX

compared participants with and without clinically significant PTSD symptom severity in terms of their improvement on the ISI ($Z$=1.17; $P$=.43), FOSQ-10 ($Z$=−1.42; $P$=.43), and PSQI ($Z$=0.76; $P$=.43) and found no statistically significant differences.

**Table 3.** Mental health symptom measures at preintervention and postintervention (N=26).

| Measures | Scale range | CBT-i[a] only group (n=14) | | | | CBT-i+physical activity group (n=12) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Preintervention | | Postintervention | | Preintervention | | Postintervention | |
| | | Values, mean (SD) | Values, median (range) | Values, mean (SD) | Values, median (range) | Values, mean (SD) | Values, median (range) | Values, mean (SD) | Values, median (range) |
| PCL-5[b] | 0-80 | 24.21 (17.55) | 20.00 (3.00-51.00) | 19.43 (13.50) | 13.50 (0.00-57.00) | 39.50 (20.21) | 42.50 (10.00-61.00) | 38.08 (22.14) | 40.00 (8.00-76.00) |
| PHQ[c]-15 | 0-30 | 9.57 (5.06) | 8.00 (4.00-22.00) | 7.86 (5.02) | 6.50 (1.00-18.00) | 11.67 (5.61) | 10.50 (5.00-21.00) | 10.42 (5.48) | 10.50 (2.00-19.00) |
| PHQ-9 | 0-27 | 9.00 (4.22) | 10.50 (2.00-14.00) | 7.21 (5.41) | 5.50 (0.00-18.00) | 11.75 (5.31) | 11.50 (5.00-22.00) | 10.67 (5.57) | 9.50 (2.00-21.00) |
| PDI[d] | 0-70 | 17.43 (12.36) | 18.50 (0.00-37.00) | 16.71 (15.86) | 11.50 (0.00-51.00) | 24.42 (13.02) | 24.50 (2.00-47.00) | 23.33 (14.42) | 25.00 (1.00-48.00) |

[a]CBT-i: cognitive behavioral therapy for insomnia coach app.

[b]PCL-5: posttraumatic stress disorder checklist–5 items.

[c]PHQ: Patient Health Questionnaire.

[d]PDI: pain disability index.

## Discussion

### Principal Findings

We conducted an RCT with 2 arms (CBT-i Coach app and CBT-i Coach app use plus PA) in veterans reporting chronic insomnia to (1) extend our previous pilot findings to a larger and more inclusive group of veterans and (2) examine the additional effect of an adjunctive PA intervention. Although PA manipulation was not effective, this study provides further evidence of the positive impact of the CBT-i Coach app for veterans with insomnia. We replicated our prior CBT-i Coach app findings [40], observing significant improvement from preintervention to postintervention in self-reported sleep quality and functional outcomes related to sleep, this time in a larger and more heterogeneous veteran sample. Participants also reported significant improvement in the ISI, although the median change of 5 points was just below the suggested clinically significant 6-point reduction [73]. Objective sleep efficiency data from the WatchPAT device also revealed a small but significant improvement in sleep efficiency. This concordance of improvement in both subjective and objective sleep measures is notable, as past research has not observed consistent results across subjective and objective sleep measures [74]. This supports the conclusion that self-management–based use of the CBT-i Coach app can improve multiple indicators of sleep quality, quantity, efficiency, and functioning.

### Insomnia and PTSD Findings

Despite randomization, veterans in the CBT-i +PA arm had significantly higher rates of PTSD. Those with PTSD symptomology that fell above a cutoff suggestive of clinical severity reported significantly worse insomnia at baseline. This association between PTSD and insomnia is partially reflected in the diagnostic criteria for PTSD, which includes nightmares and sleep disturbances [75], and thus may be because of somatic PTSD symptoms similar to or associated with symptoms of insomnia. Alternatively, veterans with PTSD often experience nightmares and heightened alertness to somatic sensations that can negatively affect their ability to stay asleep.

Interestingly, participants with and without clinical-level PTSD symptom severity had significantly and similarly improved sleep outcomes after 6 weeks of using the CBT-i Coach app. Prior work has shown that CBTI is an efficacious treatment for individuals with comorbid PTSD [59]; however, no prior work has examined this efficacy for app-delivered CBTI. This finding is of practical importance, as a recent survey revealed that some veterans with PTSD prefer treatments focusing on insomnia rather than on PTSD and may also prefer mobile or remote-delivered self-management treatment for insomnia [76]. Mobile-delivered insomnia treatments may additionally provide an entryway to interventions for those who are initially unwilling or unable to access in-person PTSD or other mental health treatments. Future work is needed to disambiguate the potential condition or symptom overlap of PTSD and insomnia and the impact of the CBT-i Coach app on comorbid insomnia and PTSD symptoms.

### Sleep Apnea and CBTI Mobile Intervention

Our results also emphasize the importance of including veterans with mild or moderate sleep apnea when evaluating behavioral interventions for insomnia. In prior work, we excluded participants with greater than mild sleep apnea, as we were concerned that the CBT-i Coach app may not be effective when the probable etiology of the sleep disturbance was OSA. Although here we excluded participants with severe sleep apnea, the broader inclusion criteria in this study led to a sample where almost one-third (10/32, 31%) had mild-to-moderate sleep apnea. These participants also benefited from the mobile



XSL•FO
RenderX

intervention, with significant improvement in insomnia severity and sleep quality at par with those without apnea. This finding suggests that a mobile CBTI intervention can still benefit veterans with mild-to-moderate sleep apnea.

## Limitations and Future Directions

Several study limitations must be considered. First, this study used a small sample (n<20 per group) and a single site. Thus, the findings may not be representative of a larger population of veterans with insomnia. Future studies with larger samples will enable the use of parametric analyses and potential covariates (eg, sex and BMI). Second, our PA intervention did not result in the needed step increase that would have enabled us to effectively test our between-group hypotheses. This manipulation failure may have been because of high preintervention steps for the PA group, which had a baseline mean step count of 9942 (SD 4841), only a few steps away from the 10,000-step goal typically given in PA interventions (ie, a ceiling effect) [77]. The CBT-i Coach app also includes some recommendations for engaging in regular exercise. Thus, it is possible that the non-PA group may have also increased their step count, although they were given self-management instructions that did not mention the PA components and did not receive the walking guide or step-tracking motivational sheet provided to the intervention group. Future research should examine the potential adjunctive role of increased PA in a group that is more sedentary and typical of young to middle-aged US samples.

Finally, because of the failure of PA manipulation, we must be cautious in interpreting outcomes related to the use of the CBT-i Coach app. Owing to our manipulation failure (ie, that participants did not increase their PA through our self-management program), our planned methodology for a pilot, single-blind, 2-group randomized controlled study lacked an effective control group. We also lacked a usual care or an additional intervention group, which would have allowed for a more robust analysis of the CBT-i Coach app's impact on sleep and functioning. Future research should include a control group that does not use the CBT-i Coach app for more robust comparison testing. We attempted to address this issue by using previously validated clinically significant improvement cutoffs for the ISI, PSQI, and FOSQ-10 to allow for more justifiable conclusions regarding the effect sizes of the within-participant sleep improvements.

## Conclusions

This study found significant improvements in subjective and objective sleep across all veterans with chronic insomnia who used mobile-delivered CBTI for 6 weeks. In a group instructed to increase their PA and that had greater PTSD symptom severity, no significant improvements in PA were observed. However, these participants instructed to increase their PA were also more physically active at baseline than expected (ie, greater baseline step counts than typical community participants) and thus had relatively less room for improvement. Importantly, we observed improved sleep outcomes in participants with clinically significant PTSD symptoms and in those with mild-to-moderate sleep apnea after app use. This suggests that interventions using mobile-delivered CBTI can be a feasible and efficacious self-management–based insomnia intervention for many veterans, including those with comorbid conditions such as PTSD and those with mild-to-moderate sleep apnea.

## Acknowledgments

This work was supported by a pilot grant to KSQ, under the Boston Roybal Center for Active Lifestyle Interventions grant (principal investigator: Margie Lachman; grant number P30 AG048785) supported by the National Institute on Aging, the Department of Veterans Affairs VISN1 Clinical Trials Network, and the Veterans Affairs Rehabilitation Research and Development-funded Center for Social and Community Reintegration Research (principal investigator: KSQ). The findings and interpretations of the data expressed in this paper are the sole responsibility of the authors and do not necessarily represent the views of the Department of Veterans Affairs. This pilot trial was registered at ClinicalTrials.gov (NCT03305354).

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile self-management guide for cognitive behavioral therapy for insomnia+physical activity group.
[PDF File (Adobe PDF File), 275 KB - formative_v5i12e29573_app1.pdf ]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2984 KB - formative_v5i12e29573_app2.pdf ]

## References

1. McLay R, Klam WP, Volkert SL. Insomnia is the most commonly reported symptom and predicts other symptoms of post-traumatic stress disorder in U.S. service members returning from military deployments. Military Medicine 2010 Oct;175(10):759-762 [FREE Full text] [doi: 10.7205/milmed-d-10-00193] [Medline: 20968266]



XSL•FO
RenderX

2. Hoge CW, McGurk D, Thomas JL, Cox AL, Engel CC, Castro CA. Mild traumatic brain injury in U.S. Soldiers returning from Iraq. N Engl J Med 2008 Jan 31;358(5):453-463. [doi: 10.1056/NEJMoa072972] [Medline: 18234750]
3. Germain A, James J, Insana S, Herringa RJ, Mammen O, Price J, et al. A window into the invisible wound of war: functional neuroimaging of REM sleep in returning combat veterans with PTSD. Psychiatry Res 2013 Feb 28;211(2):176-179 [FREE Full text] [doi: 10.1016/j.pscychresns.2012.05.007] [Medline: 23149024]
4. Lew HL, Pogoda TK, Hsu PT, Cohen S, Amick MM, Baker E, et al. Impact of the "polytrauma clinical triad" on sleep disturbance in a department of veterans affairs outpatient rehabilitation setting. Am J Phys Med Rehabil 2010 Jun;89(6):437-445. [doi: 10.1097/PHM.0b013e3181ddd301] [Medline: 20489391]
5. Alexander M, Ray MA, Hébert JR, Youngstedt SD, Zhang H, Steck SE, et al. The National Veteran Sleep Disorder study: descriptive epidemiology and secular trends, 2000-2010. Sleep 2016 Jul 01;39(7):1399-1410 [FREE Full text] [doi: 10.5665/sleep.5972] [Medline: 27091538]
6. Lindsay DR, Dyche J. Sleep disturbance implications for modern military operations. Hum Perf Extrem Environ 2012 Oct 11;10(1). [doi: 10.7771/2327-2937.1001]
7. Troxel WM, Shih RA, Pedersen ER, Geyer L, Fisher MP, Griffin BA, et al. Sleep in the military: promoting healthy sleep among U.S. Servicemembers. Rand Health Q 2015 Nov 30;5(2):19 [FREE Full text] [Medline: 28083395]
8. Spelman JF, Hunt SC, Seal KH, Burgo-Black AL. Post deployment care for returning combat veterans. J Gen Intern Med 2012 Sep;27(9):1200-1209 [FREE Full text] [doi: 10.1007/s11606-012-2061-1] [Medline: 22648608]
9. Pigeon WR, Campbell CE, Possemato K, Ouimette P. Longitudinal relationships of insomnia, nightmares, and PTSD severity in recent combat veterans. J Psychosom Res 2013 Dec;75(6):546-550. [doi: 10.1016/j.jpsychores.2013.09.004] [Medline: 24290044]
10. Wright KM, Britt TW, Bliese PD, Adler AB, Picchioni D, Moore D. Insomnia as predictor versus outcome of PTSD and depression among Iraq combat veterans. J Clin Psychol 2011 Dec;67(12):1240-1258 [FREE Full text] [doi: 10.1002/jclp.20845] [Medline: 22065464]
11. Balba NM, Elliott JE, Weymann KB, Opel RA, Duke JW, Oken BS, et al. Increased sleep disturbances and pain in veterans with comorbid traumatic brain injury and posttraumatic stress disorder. J Clin Sleep Med 2018 Nov 15;14(11):1865-1878 [FREE Full text] [doi: 10.5664/jcsm.7482] [Medline: 30373686]
12. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19 [FREE Full text] [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
13. Riemann D, Voderholzer U. Primary insomnia: a risk factor to develop depression? J Affect Disord 2003 Sep;76(1-3):255-259. [doi: 10.1016/s0165-0327(02)00072-1] [Medline: 12943956]
14. Vargas I, Perlis ML. Insomnia and depression: clinical associations and possible mechanistic links. Curr Opin Psychol 2020 Aug;34:95-99. [doi: 10.1016/j.copsyc.2019.11.004] [Medline: 31846870]
15. Kleckner IR, Zhang J, Touroutoglou A, Chanes L, Xia C, Simmons WK, et al. Evidence for a large-scale brain system supporting allostasis and interoception in humans. Nat Hum Behav 2017;1:0069 [FREE Full text] [doi: 10.1038/s41562-017-0069] [Medline: 28983518]
16. Barrett LF, Quigley KS, Hamilton P. An active inference theory of allostasis and interoception in depression. Philos Trans R Soc Lond B Biol Sci 2016 Nov 19;371(1708):20160011 [FREE Full text] [doi: 10.1098/rstb.2016.0011] [Medline: 28080969]
17. McAndrew LM, D'Andrea E, Lu SE, Abbi B, Yan GW, Engel C, et al. What pre-deployment and early post-deployment factors predict health function after combat deployment?: a prospective longitudinal study of Operation Enduring Freedom (OEF)/Operation Iraqi Freedom (OIF) soldiers. Health Qual Life Outcomes 2013 Apr 30;11:73 [FREE Full text] [doi: 10.1186/1477-7525-11-73] [Medline: 23631419]
18. Sayer NA, Noorbaloochi S, Frazier P, Carlson K, Gravely A, Murdoch M. Reintegration problems and treatment interests among Iraq and Afghanistan combat veterans receiving VA medical care. Psychiatr Serv 2010 Jun;61(6):589-597. [doi: 10.1176/ps.2010.61.6.589] [Medline: 20513682]
19. Quigley KS, McAndrew LM, Almeida L, D'Andrea EA, Engel CC, Hamtil H, et al. Prevalence of environmental and other military exposure concerns in Operation Enduring Freedom and Operation Iraqi Freedom veterans. J Occup Environ Med 2012 Jun;54(6):659-664. [doi: 10.1097/JOM.0b013e3182570506] [Medline: 22588478]
20. Pigeon WR, Britton PC, Ilgen MA, Chapman B, Conner KR. Sleep disturbance preceding suicide among veterans. Am J Public Health 2012 Mar;102 Suppl 1:S93-S97. [doi: 10.2105/AJPH.2011.300470] [Medline: 22390611]
21. Sayer NA, Frazier P, Orazem RJ, Murdoch M, Gravely A, Carlson KF, et al. Military to civilian questionnaire: a measure of postdeployment community reintegration difficulty among veterans using Department of Veterans Affairs medical care. J Trauma Stress 2011 Dec;24(6):660-670. [doi: 10.1002/jts.20706] [Medline: 22162082]
22. Hinkson Jr JK, Drake-Brooks MM, Christensen KL, Chatterley MD, Robinson AK, Crowell SE, et al. An examination of the mental health and academic performance of student veterans. J Am Coll Health 2021 Jan 05:1-8. [doi: 10.1080/07448481.2020.1858837] [Medline: 33400638]
23. Hammer LB, Brady JM, Perry ML. Training supervisors to support veterans at work: effects on supervisor attitudes and employee sleep and stress. J Occup Organ Psychol 2019 Dec 11;93(2):273-301 [FREE Full text] [doi: 10.1111/joop.12299]




XSL•FO
RenderX

24. Hom MA, Chu C, Schneider ME, Lim IC, Hirsch JK, Gutierrez PM, et al. Thwarted belongingness as an explanatory link between insomnia symptoms and suicidal ideation: findings from three samples of military service members and veterans. J Affect Disord 2017 Feb;209:114-123 [FREE Full text] [doi: 10.1016/j.jad.2016.11.032] [Medline: 27898373]
25. Bovin MJ, Miller CJ, Koenig CJ, Lipschitz JM, Zamora KA, Wright PB, et al. Veterans' experiences initiating VA-based mental health care. Psychol Serv 2019 Nov;16(4):612-620 [FREE Full text] [doi: 10.1037/ser0000233] [Medline: 29781656]
26. Possemato K, Wray LO, Johnson E, Webster B, Beehler GP. Facilitators and barriers to seeking mental health care among primary care veterans with posttraumatic stress disorder. J Trauma Stress 2018 Oct;31(5):742-752. [doi: 10.1002/jts.22327] [Medline: 30338576]
27. Gellis L, Gehrman P. Cognitive behavioral treatment for insomnia in veterans with long-standing posttraumatic stress disorder: a pilot study. J Aggress Maltreat Trauma 2011 Nov;20(8):904-916 [FREE Full text] [doi: 10.1080/10926771.2011.613448]
28. Pigeon WR, Bishop TM, Krueger KM. Insomnia as a precipitating factor in new onset mental illness: a systematic review of recent findings. Curr Psychiatry Rep 2017 Aug;19(8):44. [doi: 10.1007/s11920-017-0802-x] [Medline: 28616860]
29. Ford ES, Wheaton AG, Cunningham TJ, Giles WH, Chapman DP, Croft JB. Trends in outpatient visits for insomnia, sleep apnea, and prescriptions for sleep medications among US adults: findings from the National Ambulatory Medical Care survey 1999-2010. Sleep 2014 Aug 01;37(8):1283-1293 [FREE Full text] [doi: 10.5665/sleep.3914] [Medline: 25083008]
30. Mitchell MD, Gehrman P, Perlis M, Umscheid CA. Comparative effectiveness of cognitive behavioral therapy for insomnia: a systematic review. BMC Fam Pract 2012 May 25;13:40 [FREE Full text] [doi: 10.1186/1471-2296-13-40] [Medline: 22631616]
31. Miller MB, Donahue ML, Carey KB, Scott-Sheldon LA. Insomnia treatment in the context of alcohol use disorder: a systematic review and meta-analysis. Drug Alcohol Depend 2017 Dec 01;181:200-207 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.09.029] [Medline: 29096290]
32. NIH State-of-the-Science Conference Statement on manifestations and management of chronic insomnia in adults. NIH Consens State Sci Statements 2005;22(2):1-30 [FREE Full text] [Medline: 17308547]
33. Boness CL, Hershenberg R, Kaye J, Mackintosh M, Grasso DJ, Noser A, et al. An evaluation of cognitive behavioral therapy for insomnia: a systematic review and application of Tolin's criteria for empirically supported treatments. Clin Psychol (New York) 2020 Dec;27(4):e12348. [doi: 10.1111/cpsp.12348] [Medline: 33692609]
34. Trauer J, Qian MY, Doyle JS, Rajaratnam SM, Cunnington D. Cognitive behavioral therapy for chronic insomnia: a systematic review and meta-analysis. Ann Intern Med 2015 Aug 04;163(3):191-204. [doi: 10.7326/M14-2841] [Medline: 26054060]
35. Harb GC, Cook JM, Gehrman PR, Gamble GM, Ross RJ. Post-traumatic stress disorder nightmares and sleep disturbance in Iraq war veterans: a feasible and promising treatment combination. J Aggression Maltreatment Trauma 2009 Jul 23;18(5):516-531 [FREE Full text] [doi: 10.1080/10926770903035150]
36. Trockel M, Karlin BE, Taylor CB, Manber R. Cognitive behavioral therapy for insomnia with veterans: evaluation of effectiveness and correlates of treatment outcomes. Behav Res Ther 2014 Feb;53:41-46. [doi: 10.1016/j.brat.2013.11.006] [Medline: 24412462]
37. Epstein DR, Babcock-Parziale JL, Herb CA, Goren K, Bushnell ML. Feasibility test of preference-based insomnia treatment for Iraq and Afghanistan war veterans. Rehabil Nurs 2013;38(3):120-132. [doi: 10.1002/rnj.67] [Medline: 23319338]
38. Clough BA, Casey LM. Technological adjuncts to increase adherence to therapy: a review. Clin Psychol Rev 2011 Jul;31(5):697-710. [doi: 10.1016/j.cpr.2011.03.006] [Medline: 21497153]
39. Kang SG, Kang JM, Cho SJ, Ko KP, Lee YJ, Lee HJ, et al. Cognitive behavioral therapy using a mobile application synchronizable with wearable devices for insomnia treatment: a pilot study. J Clin Sleep Med 2017 Apr 15;13(4):633-640 [FREE Full text] [doi: 10.5664/jcsm.6564] [Medline: 28162145]
40. Reilly ED, Robinson SA, Petrakis BA, Kuhn E, Pigeon WR, Wiener RS, et al. Mobile app use for insomnia self-management: pilot findings on sleep outcomes in veterans. Interact J Med Res 2019 Jul 24;8(3):e12408 [FREE Full text] [doi: 10.2196/12408] [Medline: 31342904]
41. Kuhn E, Weiss BJ, Taylor KL, Hoffman JE, Ramsey KM, Manber R, et al. CBT-I coach: a description and clinician perceptions of a mobile app for cognitive behavioral therapy for insomnia. J Clin Sleep Med 2016 Apr 15;12(4):597-606 [FREE Full text] [doi: 10.5664/jcsm.5700] [Medline: 26888586]
42. Koffel E, Kuhn E, Petsoulis N, Erbes CR, Anders S, Hoffman JE, et al. A randomized controlled pilot study of CBT-I Coach: feasibility, acceptability, and potential impact of a mobile phone application for patients in cognitive behavioral therapy for insomnia. Health Informatics J 2018 Mar;24(1):3-13 [FREE Full text] [doi: 10.1177/1460458216656472] [Medline: 27354394]
43. Babson KA, Ramo DE, Baldini L, Vandrey R, Bonn-Miller MO. Mobile app-delivered cognitive behavioral therapy for insomnia: feasibility and initial efficacy among veterans with cannabis use disorders. JMIR Res Protoc 2015 Jul 17;4(3):e87 [FREE Full text] [doi: 10.2196/resprot.3852] [Medline: 26187404]
44. Sullivan Bisson AN, Robinson SA, Lachman ME. Walk to a better night of sleep: testing the relationship between physical activity and sleep. Sleep Health 2019 Oct;5(5):487-494 [FREE Full text] [doi: 10.1016/j.sleh.2019.06.003] [Medline: 31358470]




XSL•FO
RenderX

45. Hartescu I, Morgan K, Stevinson CD. Increased physical activity improves sleep and mood outcomes in inactive people with insomnia: a randomized controlled trial. J Sleep Res 2015 Oct;24(5):526-534 [FREE Full text] [doi: 10.1111/jsr.12297] [Medline: 25903450]
46. Passos GS, Poyares D, Santana MG, Rodrigues D'Aurea CV, Youngstedt SD, Tufik S, et al. Effects of moderate aerobic exercise training on chronic primary insomnia. Sleep Med 2011 Dec;12(10):1018-1027. [doi: 10.1016/j.sleep.2011.02.007] [Medline: 22019457]
47. Dias de Andrade FM, Pedrosa RP. The role of physical exercise in obstructive sleep apnea. J Bras Pneumol 2016;42(6):457-464 [FREE Full text] [doi: 10.1590/S1806-37562016000000156] [Medline: 28117479]
48. Atoui S, Chevance G, Romain AJ, Kingsbury C, Lachance JP, Bernard P. Daily associations between sleep and physical activity: a systematic review and meta-analysis. Sleep Med Rev 2021 Jun;57:101426. [doi: 10.1016/j.smrv.2021.101426] [Medline: 33571893]
49. Yalamanchali S, Farajian V, Hamilton C, Pott TR, Samuelson CG, Friedman M. Diagnosis of obstructive sleep apnea by peripheral arterial tonometry: meta-analysis. JAMA Otolaryngol Head Neck Surg 2013 Dec;139(12):1343-1350. [doi: 10.1001/jamaoto.2013.5338] [Medline: 24158564]
50. Carney CE, Manber R. Quiet Your Mind & Get to Sleep: Solutions to Insomnia for Those With Depression, Anxiety, or Chronic Pain. Oakland, CA, United States: New Harbinger Publications; 2009.
51. Stepping Out Mature Adults: Be Healthy, Walk Safely [Brochure]. Washington, DC: U.S. Department of Transportation; 2003.
52. Herscovici S, Pe'er A, Papyan S, Lavie P. Detecting REM sleep from the finger: an automatic REM sleep algorithm based on peripheral arterial tone (PAT) and actigraphy. Physiol Meas 2007 Feb;28(2):129-140. [doi: 10.1088/0967-3334/28/2/002] [Medline: 17237585]
53. Dontje ML, de Groot M, Lengton RR, van der Schans CP, Krijnen WP. Measuring steps with the Fitbit activity tracker: an inter-device reliability study. J Med Eng Technol 2015;39(5):286-290. [doi: 10.3109/03091902.2015.1050125] [Medline: 26017748]
54. Morin CM. Measuring outcomes in randomized clinical trials of insomnia treatments. Sleep Med Rev 2003 Jun;7(3):263-279. [doi: 10.1053/smrv.2002.0274] [Medline: 12927124]
55. Norton M. Modified Mini-Mental State Exam--Telephone Adapted Version. APA PsycTests. 1999. URL: https://doi.org/10.1037/t49761-000 [accessed 2021-11-23]
56. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
57. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? Epidemiology 2008 Nov;19(6):838-845 [FREE Full text] [doi: 10.1097/EDE.0b013e318187a7b0] [Medline: 18854708]
58. Morin CM, Belleville G, Bélanger L, Ivers H. The insomnia severity index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
59. Ulmer CS, Edinger JD, Calhoun PS. A multi-component cognitive-behavioral intervention for sleep disturbance in veterans with PTSD: a pilot study. J Clin Sleep Med 2011 Feb 15;7(1):57-68 [FREE Full text] [Medline: 21344046]
60. Chasens ER, Ratcliffe SJ, Weaver TE. Development of the FOSQ-10: a short version of the Functional Outcomes of Sleep Questionnaire. Sleep 2009 Jul;32(7):915-919 [FREE Full text] [doi: 10.1093/sleep/32.7.915] [Medline: 19639754]
61. Lubben JE. Assessing social networks among elderly populations. Fam Community Health 1988;11(3):42-52 [FREE Full text] [doi: 10.1097/00003727-198811000-00008]
62. Willer B, Rosenthal M, Kreutzer JS, Gordon WA, Rempel R. Assessment of community integration following rehabilitation for traumatic brain injury. J Head Trauma Rehabil 1993;8(2):75-87 [FREE Full text] [doi: 10.1097/00001199-199308020-00009]
63. Selim AJ, Rogers W, Fleishman JA, Qian SX, Fincke BG, Rothendler JA, et al. Updated U.S. population standard for the Veterans RAND 12-item health survey (VR-12). Qual Life Res 2009 Feb;18(1):43-52. [doi: 10.1007/s11136-008-9418-2] [Medline: 19051059]
64. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [doi: 10.1097/00006842-200203000-00008] [Medline: 11914441]
65. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002 Sep;32(9):509-515 [FREE Full text] [doi: 10.3928/0048-5713-20020901-06]
66. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: psychometric properties. Pain 1990 Feb;40(2):171-182. [doi: 10.1016/0304-3959(90)90068-O] [Medline: 2308763]
67. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]



XSL•FO
RenderX

68. Stepping out: mature adults be healthy, walk safely. National Highway Traffic Safety Administration (NHTSA). 2003. URL: https://one.nhtsa.gov/people/injury/olddrive/steppingout/pdf_version/stepping_out.pdf [accessed 2021-11-23]
69. Agnew Jr HW, Webb WB, Williams RL. The first night effect: an EEG study of sleep. Psychophysiology 1966 Jan;2(3):263-266. [doi: 10.1111/j.1469-8986.1966.tb02650.x] [Medline: 5903579]
70. Toussaint M, Luthringer R, Schaltenbrand N, Carelli G, Lainey E, Jacqmin A, et al. First-night effect in normal subjects and psychiatric inpatients. Sleep 1995 Jul;18(6):463-469. [doi: 10.1093/sleep/18.6.463] [Medline: 7481418]
71. Erdfelder E, Faul F, Buchner A. GPOWER: a general power analysis program. Behav Res Methods Instrum Comput 1996 Mar;28(1):1-11 [FREE Full text] [doi: 10.3758/bf03203630]
72. Servantes DM, Pelcerman A, Salvetti XM, Salles A, de Albuquerque PF, Alves de Salles FC, et al. Effects of home-based exercise training for patients with chronic heart failure and sleep apnoea: a randomized comparison of two different programmes. Clin Rehabil 2012 Jan;26(1):45-57. [doi: 10.1177/0269215511403941] [Medline: 21937519]
73. Yang M, Morin CM, Schaefer K, Wallenstein GV. Interpreting score differences in the Insomnia Severity Index: using health-related outcomes to define the minimally important difference. Curr Med Res Opin 2009 Oct;25(10):2487-2494. [doi: 10.1185/03007990903167415] [Medline: 19689221]
74. Tikotzky L, Sadeh A. Sleep patterns and sleep disruptions in kindergarten children. J Clin Child Psychol 2001 Dec;30(4):581-591. [doi: 10.1207/S15374424JCCP3004_13] [Medline: 11708244]
75. Diagnostic and Statistical Manual of Mental Disorders, fifth edition. Arlington, VA: American Psychiatric Association; 2013.
76. Gutner CA, Pedersen ER, Drummond SP. Going direct to the consumer: examining treatment preferences for veterans with insomnia, PTSD, and depression. Psychiatry Res 2018 May;263:108-114 [FREE Full text] [doi: 10.1016/j.psychres.2018.02.045] [Medline: 29524908]
77. Yuenyongchaiwat K. Effects of 10,000 steps a day on physical and mental health in overweight participants in a community setting: a preliminary study. Braz J Phys Ther 2016;20(4):367-373 [FREE Full text] [doi: 10.1590/bjpt-rbf.2014.0160] [Medline: 27556393]

## Abbreviations

**AHI:** apnea–hypopnea index
**CBT-i Coach:** cognitive behavioral therapy for insomnia coach app
**CBTI:** cognitive behavioral therapy for insomnia
**CIQ:** Community Integration Questionnaire
**CR:** community reintegration
**FOSQ-10:** Functional Outcomes of Sleep Questionnaire–10 items
**ISI:** insomnia severity index
**LSNS-6:** Lubben Social Network Scale–6 items
**MCS:** mental component score
**OSA:** obstructive sleep apnea
**PA:** physical activity
**PCL-5:** posttraumatic stress disorder checklist–5
**PCS:** physical component score
**PHQ:** Patient Health Questionnaire
**PSQI:** Pittsburgh Sleep Quality Index
**PTSD:** posttraumatic stress disorder
**RCT:** randomized controlled trial
**REM:** rapid eye movement
**VR-12:** Veterans RAND 12 Item Health Survey

*Edited by G Eysenbach; submitted 12.04.21; peer-reviewed by A Baharav, R de la Vega; comments to author 30.07.21; revised version received 22.10.21; accepted 28.10.21; published 09.12.21.*

*Please cite as:*
*Reilly ED, Robinson SA, Petrakis BA, Gardner MM, Wiener RS, Castaneda-Sceppa C, Quigley KS*
*Mobile Intervention to Improve Sleep and Functional Health of Veterans With Insomnia: Randomized Controlled Trial*
*JMIR Form Res 2021;5(12):e29573*
*URL: https://formative.jmir.org/2021/12/e29573*
*doi:10.2196/29573*
*PMID:34889746*



XSL•FO
RenderX

©Erin Dawna Reilly, Stephanie A Robinson, Beth Ann Petrakis, Melissa M Gardner, Renda Soylemez Wiener, Carmen Castaneda-Sceppa, Karen S Quigley. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Tracking Lower Urinary Tract Symptoms and Tamsulosin Side Effects Among Older Men Using a Mobile App (PERSONAL): Feasibility and Usability Study

Austin W Lee[1], BA, MD; Stacey A Kenfield[1,2], BS, MS, DSc; Elizabeth Y Wang[3], BA, MD; Anthony Enriquez[1], BA; Akinyemi Oni-Orisan[4], BS, PharmD, PhD; Michael A Steinman[5,6], BA, MD; Ida Sim[7], MD, PhD; Benjamin N Breyer[1,2*], BS, MD, MAS; Scott R Bauer[1,5,7*], BA, MD, MS

[1]Department of Urology, University of California, San Francisco, San Francisco, CA, United States
[2]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, United States
[3]Columbia University Vagelos College of Physicians and Surgeons, New York City, CA, United States
[4]Department of Clinical Pharmacy and Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, United States
[5]Veterans Affairs Medical Center, University of California, San Francisco, San Francisco, CA, United States
[6]Division of Geriatrics, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States
[7]Division of General Internal Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States
[*]these authors contributed equally

**Corresponding Author:**
Scott R Bauer, BA, MD, MS
Department of Urology
University of California, San Francisco
550 16th Street, Box 1695
San Francisco, CA, 94158
United States
Phone: 1 415 221 4810 ext 24322
Email: Scott.Bauer@ucsf.edu

**Related Article:**

Companion article: https://humanfactors.jmir.org/2021/4/e30767

## Abstract

**Background:** Continuous α1a-blockade is the first-line treatment for lower urinary tract symptoms (LUTS) among older men with suspected benign prostatic hyperplasia. Variable efficacy and safety for individual men necessitate a more personalized, data-driven approach to prescribing and deprescribing tamsulosin for LUTS in older men.

**Objective:** We aim to evaluate the feasibility and usability of the PERSONAL (Placebo–Controlled, Randomized, Patient-Selected Outcomes, N-of-1 Trials) mobile app for tracking daily LUTS severity and medication side effects among older men receiving chronic tamsulosin therapy.

**Methods:** We recruited patients from the University of California, San Francisco health care system to participate in a 2-week pilot study. The primary objectives were to assess recruitment feasibility, study completion rates, frequency of symptom tracking, duration of tracking sessions, and app usability rankings measured using a follow-up survey. As secondary outcomes, we evaluated whether daily symptom tracking led to changes in LUTS severity, perceptions of tamsulosin, overall quality of life, medication adherence between baseline and follow-up surveys, and perceived app utility.

**Results:** We enrolled 19 men within 23 days, and 100% (19/19) of the participants completed the study. Each participant selected a unique combination of symptoms to track and recorded data in the PERSONAL app, with a median daily completion rate of 79% (11/14 days). The median duration of the app session was 44 (IQR 33) seconds. On a scale of 1 (strongly disagree) to 5 (strongly agree), the participants reported that the PERSONAL app was easy to use (mean 4.3, SD 1.0), that others could learn to use it quickly (mean 4.2, SD 0.9), and that they felt confident using the app (mean 4.4, SD 0.8). LUTS severity, quality of life, and medication adherence remained unchanged after the 2-week study period. Fewer men were satisfied with tamsulosin after



XSL•FO
RenderX

using the app (14/19, 74% vs 17/19, 89% at baseline), although the perceived benefit from tamsulosin remained unchanged (18/19, 95% at baseline and at follow-up). In total, 58% (11/19) of the participants agreed that the PERSONAL app could help people like them manage their urinary symptoms.

**Conclusions:** This pilot study demonstrated the high feasibility and usability of the PERSONAL mobile app to track patient-selected urinary symptoms and medication side effects among older men taking tamsulosin to manage LUTS. We observed that daily symptom monitoring had no adverse effects on the secondary outcomes. This proof-of-concept study establishes a framework for future mobile app studies, such as digital n-of-1 trials, to collect comprehensive individual-level data for personalized LUTS management in older men.

***(JMIR Form Res 2021;5(12):e30762)*** doi:10.2196/30762

**KEYWORDS**

LUTS; tamsulosin; mobile; app; mobile phone

# Introduction

## Background

Lower urinary tract symptoms (LUTS) affect more than half of men over the age of 70 years [1] and are associated with an increased risk of falls [2] and psychological distress that impairs health-related quality of life [3]. Clinical practice guidelines for the management of LUTS as a result of presumed benign prostatic hyperplasia list $\alpha_{1a}$-blockers, such as tamsulosin, as the first-line therapy [4]. Widespread adoption of $\alpha_{1a}$-blocker therapy for prolonged durations may expose patients, particularly those who are nonresponders, to unnecessary potential harms such as orthostatic hypotension and dizziness, which can result in falls [5] and polypharmacy [6].

Several considerations challenge the existing practice of continuous $\alpha_{1a}$-blocker use in older men with LUTS. First, randomized clinical trials demonstrate modest efficacy of tamsulosin and other $\alpha_{1a}$-blockers for improving LUTS, as well as a large placebo effect [7]. This limits the ability of clinicians to identify true responders to chronic tamsulosin therapy and the degree of symptomatic improvement attributable to tamsulosin alone [7,8]. In fact, several small studies have indicated no symptomatic worsening after tamsulosin discontinuation among men with chronic LUTS [9-11] or men concurrently taking a 5α-reductase inhibitor [12-15]. The landmark trials that informed societal guidelines were also conducted predominantly among healthy and younger (age <65 years) White men [7] and may not be generalizable to more racially diverse and older men with multimorbidity and polypharmacy who are most likely to be receiving chronic tamsulosin therapy [16,17]. Fewer benefits and greater harm in this population have led to recommendations against the use of $\alpha_{1a}$-blockers in older men [18,19]. However, tamsulosin may be safe and effective for a subset of older men with LUTS. A more personalized approach to identifying which older men will and will not benefit from continuing long-term tamsulosin therapy is urgently needed.

Mobile health (mHealth) apps are an emerging platform for personalized health care delivery that have thus far been underused in benign urology [20]. The ability of mHealth apps to achieve high-throughput, low-burden data collection [21] on a wide range of patient-centric outcomes could facilitate daily tracking of LUTS severity and side effects among men receiving chronic LUTS therapy. Previous efforts using mobile apps to track patient-reported outcomes have informed patient-driven care and helped patients with heart failure and depression identify and self-manage potential symptom triggers [22,23]. If men are able and willing to track their symptoms while undergoing LUTS treatment (eg, formal n-of-1 trials or self-experimentation), their data can be used to generate individualized estimates of benefits and harms of both prescribed and self-management LUTS treatments [24-30]. To reduce the burden of frequent symptom assessments, we designed the PERSONAL (Placebo–Controlled, Randomized, Patient-Selected Outcomes, N-of-1 Trials) mHealth app for collecting daily LUTS severity and medication side effect data. However, whether older men are able and willing to use a mobile app to track their LUTS and medication side effects, particularly while receiving chronic $\alpha_{1a}$-blocker therapy, represents a knowledge gap.

## Objectives

In this paper, we report the results of a 2-week pilot study using the PERSONAL mobile app to monitor daily LUTS severity and medication side effects among older men receiving chronic tamsulosin therapy. Our primary objectives are to assess the feasibility and usability of the PERSONAL app as demonstrated by time to reach full recruitment, study completion rates, frequency of symptom tracking, duration of tracking sessions, and app usability rankings. As secondary objectives, we explore whether quality of life measures, medication adherence, LUTS severity, and perception of tamsulosin changed after using the app, as well as the perceived utility of the app. We hypothesize that daily use of the PERSONAL app would not significantly impact these secondary outcomes.

# Methods

## Design and Setting

We recruited patients from the University of California, San Francisco health care system to participate in a 2-week mobile app–based study to evaluate the feasibility and usability of the PERSONAL app for tracking LUTS severity and tamsulosin side effects. This study was approved by the University of California, San Francisco Institutional Review Board (institutional review board approval: 19-28557), and all participants provided informed consent.

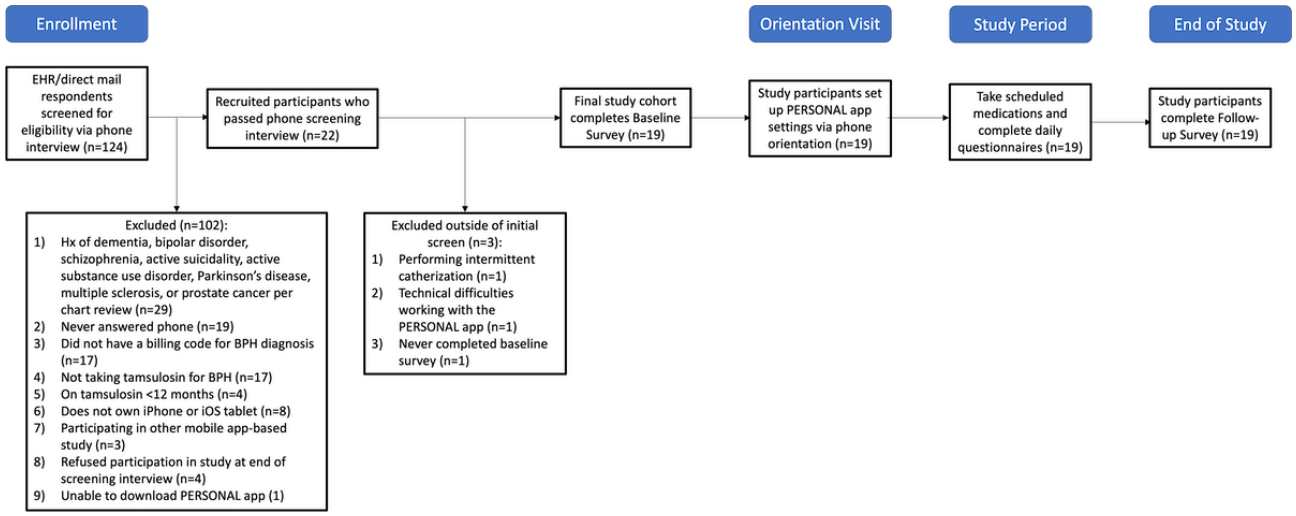


XSL•FO
RenderX

## Recruitment

We invited eligible male patients aged >55 years who were monitored by urology clinicians from the University of California, San Francisco, who had previously agreed to be contacted by researchers via the electronic medical record, and who had a diagnosis of LUTS or benign prostatic hyperplasia on the basis of International Classification of Diseases-10 billing codes (N40.x) and an active prescription for tamsulosin hydrochloride to participate in this study (Figure 1). On the basis of prior mHealth studies, we determined that we would have 90% power to observe at least one failed primary outcome during this feasibility study with a target sample size of 20 participants. Eligible patients received an invitation by secure message via the electronic medical record portal describing the PERSONAL app and study, and those who were not enrolled in secure messaging received a mailed letter with the same information. Patients who expressed interest in participating were screened for inclusion and exclusion criteria by telephone (Multimedia Appendix 1). Eligible patients were then sent a baseline questionnaire using REDCap (Research Electronic Data Capture; Vanderbilt University), instructed on how to download the PERSONAL app (available for free on the iOS App Store at the time of the study), and scheduled for a phone orientation visit with a study team member.

**Figure 1.** Study flow diagram. BPH: benign prostatic hyperplasia; EHR: electronic health record; Hx: history; PERSONAL: Placebo–Controlled, Randomized, Patient-Selected Outcomes, N-of-1 Trials.




## Orientation and Pilot

The study subjects participated in their orientation visit via a phone call with a study team member who confirmed or facilitated the installation of the PERSONAL app on the participant's iOS device. Following verbal cues from the study team member, the participants then customized the app (Figure 2A) by selecting at least 1 urinary symptom from a list of 10 and at least 1 tamsulosin side effect from a list of 12, sourced from the tamsulosin Food and Drug Administration label, to track daily for 2 weeks. The participants were informed that this list included potential tamsulosin side effects but that they might also experience them because of reasons unrelated to tamsulosin. Patient-selected symptoms and side effects were used instead of a general symptom questionnaire to minimize the daily time spent answering questions on the PERSONAL app and to maximize the utility of the app for each participant by focusing on the symptoms they found most bothersome (Figures 2B and 2D). The participants were recommended to track 2 to 3 urinary symptoms and 2 to 3 potential medication side effects, but they were allowed to track as many symptoms and side effects as they desired (Multimedia Appendix 2). They also selected the time they preferred for receiving app reminder notifications. After completing the app setup, the participants were oriented on the rest of the PERSONAL app features and practiced recording their first symptoms and side effects. At the end of the orientation visit, the participants were provided with study team contact information for technical app support as needed throughout the study period.

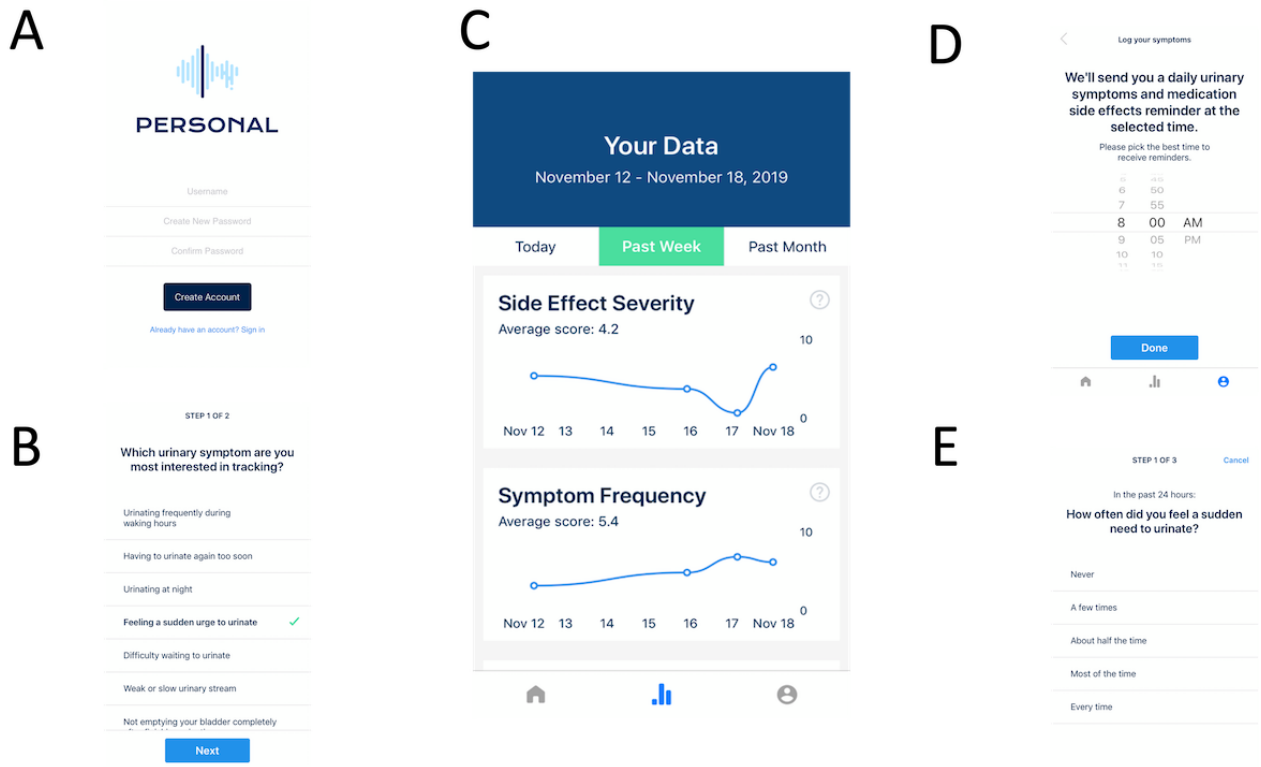
At baseline, we collected participant demographics, social and health-related behaviors, and medical history through a REDCap survey. Physical activity, alcohol use, and social connection or isolation were assessed using the Institute of Medicine Measures of Social and Behavioral Determinants of Health and categorized according to established thresholds [31]. We also assessed the perceptions of the participants regarding medication burden and appropriateness, their willingness to stop medications, and their desire to be involved in medication decisions via the revised Patients' Attitudes Toward Deprescribing (rPATD) questionnaire (not specific to tamsulosin) [32].

The participants used the PERSONAL app for 2 weeks and completed daily symptom and side effect questionnaires to track the urinary symptoms and potential tamsulosin side effects they had selected (Figure 2E). At any point during the study, the participants were able to graphically visualize their symptom and side effect severity scores over time within the PERSONAL app (Figure 2C). The total number and duration of the tracking sessions were recorded using the PERSONAL app (Multimedia Appendix 3).



XSL•FO
RenderX

**Figure 2.** Representative screenshots of the PERSONAL app interface. (A) Login. (B) Urinary symptom selection. (C) Personalized graphs of changes in symptom and side effect severity over time. (D) Daily app reminder setup. (E) Daily symptom and side effect severity questionnaire. PERSONAL: Placebo–Controlled, Randomized, Patient-Selected Outcomes, N-of-1 Trials.




## Primary Outcomes

To evaluate feasibility, we assessed the time elapsed until our recruitment goal was reached, the percentage of participants who completed the study, the mean number of days with completed symptom tracking, and the mean duration of the tracking sessions. To evaluate usability, the participants completed a previously published and validated usability scale [33] via REDCap asking them to rate qualitative statements describing the app's frequency of use, ease of use, integration of functions, and ability to be learned by new users, as well as confidence using the app. We also assessed the experience of the patients with the phone-based PERSONAL orientation visit and app setup to determine if in-person orientation visits are needed for future studies. We set the following benchmarks to measure study and app effectiveness: (1) recruitment and retention of 20 eligible men would be completed within 3 to 6 months, (2) >70% of participants would complete the study, (3) the daily questionnaire completion rate would be >50%, (4) the average symptom logging session duration would be <2 minutes, and (5) the mean rating of ease of use of the app would be >3 on a scale of 1 (strongly disagree) to 5 (strongly agree).

## Secondary Outcomes

Secondary outcomes were assessed using baseline and follow-up surveys. These included LUTS severity, perceptions of tamsulosin, overall quality of life, and medication adherence. The perceived utility of the app was also assessed at the end of the study.

LUTS severity was measured using questions selected from the validated Lower Urinary Tract Dysfunction Research Network Comprehensive Assessment of Self-Reported Urinary Symptoms questionnaire item bank, which was generated following the principles of the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative funded by the National Institutes of Health [34]. Specifically, 8 questions were asked to assess urinary urgency (frequency of urge episodes and severity of urgency), daytime frequency (number of voids and interval between voids), nocturia, slow or weak urine flow, incomplete emptying, and postvoid dribbling. The responses to these questions were added to obtain a composite LUTS severity score (range 0-30). An additional 8 questions were asked to assess urinary incontinence subtypes, including urgency, stress, and unspecified urinary incontinence. The responses were added to obtain a composite urinary incontinence severity score (range 0-32). The perceptions of tamsulosin were evaluated using questions that inquired whether the participants were satisfied with tamsulosin and whether they felt any perceived benefits from it.

We evaluated the health-related quality of life of the patients using the PROMIS-29 v2.0 [35]. PROMIS scores represent standardized T-scores for a given domain referenced to a population with mean 50 and SD 10, where a higher PROMIS T-score represents having more of the given domain (ie, a higher PROMIS T-score for anxiety represents an individual having greater anxiety). PROMIS T-score thresholds can be interpreted as follows for symptoms (ie, anxiety, depression, fatigue, sleep disturbance, and pain interference): 0 to 55, within normal limits; >55 to 60, mild; >60 to 70, moderate; and >70, severe. PROMIS T-score thresholds can be interpreted as follows for domains (ie, physical function and ability to participate in social roles):



XSL•FO
RenderX

>45, within normal limits; >40 to 45, mild deficit; >30 to 40, moderate deficit; ≤30, severe deficit.

Medication adherence was measured using the Voils Medication Adherence score [36]. The Voils Medication Adherence score ranges from 1 to 5, with higher scores indicating a greater extent of medication nonadherence. Finally, we evaluated the perceptions of the patients regarding the utility of the PERSONAL app on the basis of standardized questionnaires from previous mHealth studies [28].

Mean and SD were calculated for scores with normal distribution, and median and IQR were calculated for scores with skewed distribution. Statistical significance was set at $P<.05$. All analyses were performed using STATA (version 15.1; StataCorp LLC).

## Results

### Demographics

After our initial study invitation, 124 men expressed interest in the study, 102 were excluded via telephone screening, and 19 were ultimately enrolled in the study between March 3 and March 26, 2020 (Figure 1). The mean age of the participants was 70 (SD 7) years, and 32% (6/19) of them self-identified their race or ethnicity as Asian, Hispanic, or other (Table 1). Most participants had completed a bachelor's degree or higher (16/19, 84%) and reported no difficulty in paying basic living expenses (14/19, 74%). More than a third of the participants were socially isolated (7/19, 37% on the basis of frequency of communication or getting together with family, friends, or neighbors in a typical week; frequency of attending religious services, clubs, or organizations in the past year; and marital status), and 21% (4/19) screened positive for alcohol use. Hypertension, prostatitis, and visual impairment were the most common comorbidities, with a mean of 2 (SD 1.5) comorbidities per participant. The participants generally disagreed that medications were burdensome (mean rPATD score 2.7, SD 1.0), that their current medication regimen was inappropriate (mean rPATD score 2.7, SD 0.8), and that they had a willingness to stop medications (mean rPATD score 2.2, SD 0.8), and agreed that they desired to be involved in their medication regimen determination (mean rPATD score 4.2, SD 0.5).



XSL•FO
RenderX

**Table 1.** Demographics, health-related behaviors, and medical history of 19 study participants.

| Demographic | Value |
|---|---|
| Age (years), mean (SD) | 70 (7) |
| **Race or ethnicity, n (%)** | |
| White, non-Hispanic | 13 (68) |
| Black | 0 (0) |
| Asian | 3 (16) |
| Hispanic | 2 (11) |
| Other | 1 (5) |
| **Highest degree earned, n (%)** | |
| High school diploma | 2 (11) |
| Associate degree | 1 (5) |
| Bachelor's degree | 5 (26) |
| Master's degree | 8 (42) |
| Doctorate or professional | 3 (16) |
| **Ability to pay for basic living expenses, n (%)** | |
| Not hard at all | 14 (74) |
| Somewhat hard | 4 (21) |
| Very hard | 0 (0) |
| Prefer not to answer | 1 (5) |
| **Social and health-related behaviors, n (%)** | |
| **Physical activity**[a] | |
| Inactive | 3 (16) |
| Insufficiently active | 3 (16) |
| Sufficiently active | 13 (68) |
| Ever smoked at least 100 cigarettes | 11 (58) |
| Current smoking | 0 (0) |
| Positive screen for alcohol use[b] | 4 (21) |
| **Social connection or isolation score**[c] | |
| Most isolated | 7 (37) |
| Very isolated | 1 (5) |
| Somewhat isolated | 8 (42) |
| Not isolated | 3 (16) |
| **Self-reported medical history** | |
| **Comorbidities, n (%)** | |
| Hypertension | 10 (53) |
| Coronary artery disease | 5 (26) |
| Angina | 2 (11) |
| Congestive heart failure | 1 (5) |
| Chronic obstructive pulmonary disease | 0 (0) |
| Diabetes | 2 (11) |
| Stroke or intracerebral hemorrhage | 2 (11) |
| Parkinson disease or multiple sclerosis | 0 (0) |



XSL•FO
RenderX

| Demographic | Value |
|---|---|
| Visual impairment | 6 (32) |
| Prostatitis | 9 (47) |
| Number of comorbidities, mean (SD) | 2 (1.5) |
| **Revised Patients' Attitudes Toward Deprescribing score, mean (SD)** | |
| Burden subscore | 2.7 (1.0) |
| Appropriateness subscore | 2.7 (0.8) |
| Willingness to stop medications subscore | 2.2 (0.8) |
| Involvement in medications subscore | 4.2 (0.5) |

[a]Physical activity was calculated in minutes per week engaged in moderate to strenuous activity and was categorized as *inactive* (0 min/week), *insufficiently active* (1–149 min/week), and *sufficiently active* (150+ min/week).

[b]Alcohol use was tabulated as a composite value integrating alcohol consumption frequency (*How often do you have a drink?*) and density (*How many standard drinks on a typical day? How often do you have ≥6 drinks on one occasion?*), and a score of ≥4 indicated a positive screening.

[c]Social isolation was assessed as a composite value integrating not interacting with others, not attending social gatherings (church, meetings, or clubs), and not being married. The values could then be interpreted as follows: *most isolated* (0 to 1 point), *very isolated* (2 points), *somewhat isolated* (3 points), and *not isolated* (4 points).

## Feasibility

The time required to reach our enrollment goal was 23 days. A total of 95% (19/20) of the enrolled participants completed the study (1 participant was enrolled but did not complete the baseline survey and was therefore excluded from subsequent analyses) and tracked their symptoms on the PERSONAL app for a median of 11 out of 14 (IQR 3) days. They spent a median of 44 (IQR 33) seconds on each tracking session. When symptom severity scores were visualized over time, we observed significant heterogeneity among the participants, with some men exhibiting highly variable symptom severity and others reporting stable symptom severity (Figure 3).



XSL•FO
RenderX

**Figure 3.** Change in urinary symptom severity over time for 19 study participants. Each panel in this figure visualizes the change in urinary symptom severity for an individual study participant.

## Usability

The PERSONAL app usability results are presented in Table 2. On a scale of 1 (strongly disagree) to 5 (strongly agree), the participants generally agreed that the PERSONAL app was easy to use (mean 4.3, SD 1.0), that they could imagine people learning to use it quickly (mean 4.2, SD 0.9), and that they felt confident using the app (mean 4.4, SD 0.8). They agreed less on whether they would use the app frequently (mean 2.7, SD 1.2). The participants generally disagreed with statements describing the PERSONAL app as awkward to use, having a high learning curve, or requiring technical support to use. They agreed that the phone-based orientation was useful.



XSL•FO
RenderX

**Table 2.** Perceived usability of PERSONAL (Placebo–Controlled, Randomized, Patient-Selected Outcomes, N-of-1 Trials) app.

| Experience[a] | Value, mean (SD) |
|---|---|
| **PERSONAL app** | |
| **Pros** | |
| Would use app frequently | 2.7 (1.2) |
| Easy to use | 4.3 (1.0) |
| Various functions well integrated | 3.5 (1.2) |
| Imagine people would learn to use quickly | 4.2 (0.9) |
| Felt confident using app | 4.4 (0.8) |
| **Cons** | |
| Found app unnecessarily complex | 1.7 (0.9) |
| Would need technical support to use | 1.5 (0.9) |
| Too much inconsistency in app | 1.9 (0.8) |
| Awkward to use | 1.6 (0.9) |
| High learning curve | 1.5 (0.9) |
| **PERSONAL orientation visit and app setup** | |
| **Pros** | |
| Comprehensive to get started with study | 4.3 (0.8) |
| Tracked symptoms most important to patient | 4.2 (0.7) |
| **Cons** | |
| Could set up app without orientation | 3.6 (1.3) |
| In-person orientation would be more helpful | 1.5 (0.8) |
| Needed more guidance after | 1.7 (1.0) |
| Unable to track all prioritized symptoms | 2.7 (1.1) |
| Unable to track all prioritized side effects | 2.1 (1.0) |

[a]Scale: 1, *strongly disagree*, to 5, *strongly agree*.

## Secondary Outcomes

Table 3 displays the results of the secondary outcomes assessed at baseline and follow-up. LUTS severity remained unchanged throughout the study period (LUTS severity score at baseline: mean 12, SD 5; at follow-up: mean 12, SD 5). After using the PERSONAL app for 2 weeks, fewer men were satisfied with tamsulosin (14/19, 89% of the participants responded *Yes* when asked "*Taking all things into account, are you satisfied with your tamsulosin medication?*" vs 17/19, 74% at baseline), although the perceived benefits from tamsulosin remained unchanged (18/19, 95% responded *Yes* when asked "*Have you had any benefit from your tamsulosin medication?*" at both baseline and follow-up). Overall, the participants reported normal health-related quality of life at baseline, and their PROMIS T-scores did not worsen after using the PERSONAL app for 2 weeks. The participants also reported high medication adherence before and after using the PERSONAL app (Voils score at baseline: mean 2.3, SD 0.2; at follow-up: mean 2.4, SD 0.3).



XSL•FO
RenderX

**Table 3.** Patient-reported secondary outcomes at baseline and follow-up.

| Secondary outcome | Baseline | Follow-up |
|---|---|---|
| **LUTS[a] severity and treatment satisfaction** | | |
| LUTS severity score[b], mean (SD) | 12 (5) | 12 (5) |
| Urinary incontinence severity score[c], mean (SD) | 1 (2) | 1 (1) |
| Extremely or very bothered by urinary symptoms, n (%) | 3 (16) | 2 (11) |
| Satisfied with tamsulosin, n (%) | 17 (89) | 14 (74) |
| Any perceived benefit from tamsulosin, n (%) | 18 (95) | 18 (95) |
| **Quality of life** | | |
| **PROMIS[d]-29 v2.0 T-scores, mean (SD)** | | |
| Physical function | 52 (8) | 52 (8) |
| Anxiety | 52 (10) | 51 (9) |
| Depression | 48 (8) | 46 (7) |
| Fatigue | 47 (10) | 47 (9) |
| Sleep disturbance | 56 (5) | 54 (4) |
| Ability to participate in social roles | 56 (10) | 52 (10) |
| Pain interference | 50 (11) | 50 (9) |
| **Medication adherence** | | |
| Voils Medication Adherence score, mean (SD) | 2.3 (0.2) | 2.4 (0.3) |

[a]LUTS: lower urinary tract symptoms.

[b]Sum of nonincontinence items from the Lower Urinary Tract Dysfunction Research Network assessing urgency, daytime frequency, nocturia, slow or weak urine flow, incomplete emptying, and postvoid dribbling (range 0-30; higher score indicates greater severity).

[c]Sum of incontinence items from the Lower Urinary Tract Dysfunction Research Network assessing urgency, stress, and unspecified urinary incontinence (range 0-32; higher score indicates greater severity).

[d]PROMIS: Patient-Reported Outcomes Measurement Information System.

## Utility

The perceived utility of the PERSONAL app is presented in Table 4. More than half (11/19, 58%) of the participants reported that the PERSONAL app could help people like them manage their urinary symptoms. In the absence of additional guidance or clinician involvement, the participants rated the utility of the PERSONAL app on a scale of 1 (not at all helpful) to 5 (extremely helpful) for keeping track of symptoms (mean 2.9, SD 1.6), working with physicians to achieve treatment goals (mean 3.0, SD 1.6), noticing things that help with urinary symptoms (mean 2.6, SD 1.6), and building confidence in the approach to managing urinary symptoms (mean 2.9, SD 1.7).

**Table 4.** Perceived utility of PERSONAL (Placebo–Controlled, Randomized, Patient-Selected Outcomes, N-of-1 Trials) app.

| Item | Value |
|---|---|
| Belief that PERSONAL app can help manage LUTS[a], n (%) | 11 (58) |
| **Perceived utility of PERSONAL app for LUTS management[b], mean (SD)** | |
| Keeping track of symptoms | 2.9 (1.6) |
| Working with physicians to achieve treatment goals | 3.0 (1.6) |
| Identifying urinary symptom triggers | 2.3 (1.5) |
| Noticing things that help with urinary symptoms | 2.6 (1.6) |
| Confidence in approach to urinary symptoms | 2.9 (1.7) |

[a]LUTS: lower urinary tract symptoms.

[b]Scale: 1, *not at all helpful* to 5, *extremely helpful*.



XSL•FO
RenderX

## Discussion

### Principal Findings

In this pilot study of a mobile app for tracking daily LUTS severity and medication side effects, we demonstrated that recruiting and retaining older men receiving chronic tamsulosin therapy to use the PERSONAL app is highly feasible; the participants were able to complete the study and track their symptoms almost every day for 2 weeks, and more than half of them found the app helpful despite no clinician involvement or additional guidance beyond symptom tracking. We also observed no adverse effects of daily symptom monitoring on the secondary outcomes of LUTS severity, health-related quality of life, and medication adherence. Interestingly, fewer participants reported satisfaction with tamsulosin after the symptom tracking period.

### Feasibility of Tracking LUTS With the PERSONAL App

As a proof-of-concept study, the PERSONAL pilot demonstrated that a mobile app could be used to track daily symptoms and side effects among older men taking tamsulosin for LUTS. The participants agreed that the app was easy to use, that people could learn to use it quickly, and that they felt confident using the app without additional guidance. High observed rates of daily questionnaire completion suggest that the interface of the app and questionnaire length were not so burdensome as to discourage the participants from tracking their symptoms. The duration of the average tracking session further showed how quickly men could record their daily symptoms. The findings that the PERSONAL app can build high user engagement in a feasible length of time align with those of similar studies assessing mobile app feasibility in optimizing psychotherapy in community psychiatric clinics and routine follow-up care in breast cancer survivors [37,38]. Although longer follow-up is needed to determine how the duration of tracking influences sustained engagement and perceived burden, these findings indicate that it is possible to collect a large volume of data on urinary symptoms and medication side effects from older men with relatively low perceived burden for a period of at least 2 weeks. High-volume individualized data may be particularly valuable in clinical or research settings where changes (or lack thereof) in symptoms or side effects observed after an intervention or modification of treatment strategy could inform the patient and their clinician and guide subsequent management decisions, although this hypothesis needs to be tested in future studies.

### Heterogeneity in Patient-Selected Symptom Tracking

Of note was the degree of heterogeneity in both the symptoms prioritized for tracking among the participants and daily symptom severity during the pilot. LUTS can refer to a range of different symptoms, including both voiding sequelae such as hesitancy, poor stream, straining, prolonged urination, incomplete bladder emptying, and dribbling, and storage sequelae such as frequency, urgency incontinence, and nocturia [39]. Consequently, follow-up studies have identified different urinary symptoms as the most bothersome, among them urge urinary incontinence [40,41] and nocturia [42]. The reasons for this heterogeneity include underlying anatomical pathophysiology, personal lifestyle, cultural backgrounds, and psychosocial factors. Regardless, the capacity to capture such heterogeneity at the individual level allows for the directed management of the particular symptoms of a patient rather than a broader diagnosis. Similarly, variability in symptom severity hinges on factors such as cultural mindsets toward health and a subjective threshold to report symptoms [43], further emphasizing the need for more granular data and individualized recommendations.

### Monitoring for Unintended Consequences of Daily Symptom Tracking

After tracking their symptoms for 2 weeks using the PERSONAL app, the participants did not experience changes in LUTS severity, quality of life, or medication adherence. This may be because such perceptions of LUTS, quality of life, and medication adherence are more fixed opinions rather than perspectives that easily fluctuate, especially during a short-term study. It is also possible that, because this was a single-arm pilot without a comparator, the participants were not able to experience what life off tamsulosin might be like—which could have subsequent implications for the perception of LUTS, quality of life, and medication adherence. Although short-term studies with daily symptom tracking appear unlikely to have unintended adverse consequences, the risk may increase with studies of longer duration. Despite no change in perceptions of adherence, fewer participants were satisfied with tamsulosin after using the PERSONAL app for 2 weeks. One potential reason for this finding is that tracking their symptoms and side effects showed the participants direct evidence that their most bothersome symptoms were poorly controlled or that they experienced side effects more often than they expected. Increased monitoring may have also made them aware that their symptoms did not necessarily correlate with the dosing schedule of tamsulosin, thus diminishing its benefit. The app could also have served as a physical reminder of the burden of taking daily medication, resulting in less satisfaction with tamsulosin itself. Although our sample size was small for this pilot study, these findings serve as testable hypotheses for further exploration.

### Perceived Utility of the PERSONAL App

More than half of the participants agreed that the PERSONAL app would be helpful when working with a clinician to achieve their treatment goals. They believed that these additional data on the day-to-day variability in LUTS severity and medication side effects might influence LUTS management and could be useful to share with their clinician. However, in the absence of more detailed guidance or clinician involvement, the participants found the PERSONAL app neither helpful nor unhelpful. This finding suggests that men do require additional support to interpret the variability in their symptoms and connect this variability to various symptom triggers or treatments. In the absence of self-experimentation or alternating treatments, the PERSONAL app does not help participants determine if either their self-management techniques or tamsulosin is effective in reducing LUTS severity or side effects. This information will be used to develop more personalized recommendations for



XSL•FO
RenderX

participants in larger planned individualized crossover (or n-of-1) trials using the PERSONAL app.

### Strengths and Limitations

The strengths of our study include the demonstration of successful remote patient recruitment and enrollment, high patient engagement and participation, and the novel use of patient-selected outcomes to minimize participant burden and maximize the number of repeated symptom assessments. This study also provides key feasibility and usability data to guide improvements in the PERSONAL app. This low-burden approach to data collection using the PERSONAL app may also facilitate the use of individualized crossover or n-of-1 trials [44] to further personalize LUTS management by testing interventions with short wash-out periods, such as tamsulosin therapy, on an individual patient. Subsequent integration of self-experimentation or n-of-1 trial protocols into a mobile app would allow patients to assess their personalized response to specific therapies in terms of LUTS severity and intervention side effects [45].

Our study has some limitations. First, although one-third of the study participants were people of color, we did not recruit any Black or African American participants and cannot comment on whether the PERSONAL app has similar feasibility or usability outcomes in this population. Future studies will require more explicit recruitment goals and customized approaches to recruit a more diverse patient population, including transcultural adaptations of questionnaires and surveys. Second, the participants in our pilot study were highly educated and had a high socioeconomic status, although 21% (4/19) found it somewhat difficult to pay for basic living expenses. This limits our ability to comment on whether other populations of older men who may not be as educated or wealthy may have greater difficulty using an mHealth app to track their symptoms and medication side effects. However, other studies have demonstrated widespread accessibility of and comfort with smartphone use irrespective of socioeconomic status [46]. Third, our pilot study used detailed baseline and follow-up questionnaires that may have impacted the desire of the participants to enroll in the study. In future studies, it may be possible to reduce the number of question items while still capturing appropriate granularity, particularly for questionnaires that participants will complete repeatedly during study follow-up. Fourth, we used a version of the rPATD that assesses general attitudes toward deprescribing rather than a more recently developed version that is specific to tamsulosin [47]. However, we found that men in both studies reported similar attitudes regarding willingness to consider targeted deprescribing. Finally, the participants remained on the same LUTS treatment throughout the study period; therefore, we were unable to provide them with personalized estimates of tamsulosin efficacy or side effects, which is a direction for future study.

### Conclusions

In conclusion, the PERSONAL app pilot demonstrated high feasibility and usability of a mobile app for tracking and visualizing daily LUTS severity and medication side effects among older men receiving chronic tamsulosin therapy. The PERSONAL app did not change the LUTS severity, health-related quality of life, or medication adherence of the patients. Future directions of study include leveraging the high-volume data collected via the PERSONAL app to develop individualized estimates of benefits and harms from LUTS interventions in older men.

### Acknowledgments

This study was supported by grants to SRB from the National Institute of Diabetes, Digestive, and Kidney Disorders (grant 1K12DK111028) and the National Institute on Aging (grant 1R03AG067937), to SAK from the Claude D Pepper Older Americans Independence Center at the University of California, San Francisco funded by the National Institute on Aging (grant P30 AG044281) for the Helen Diller Family Chair in Population Science for Urologic Cancer at the University of California, San Francisco, and to MAS from the National Institute on Aging (grants K24AG049057 and R24AG064025).

### Conflicts of Interest

None declared.

Multimedia Appendix 1
Study inclusion and exclusion criteria.
[DOC File , 30 KB - formative_v5i12e30762_app1.doc ]

Multimedia Appendix 2
Participants' ranking of most bothersome lower urinary tract symptoms and tamsulosin side effects to track using the PERSONAL (Placebo–Controlled, Randomized, Patient-Selected Outcomes, N-of-1 Trials) app.
[DOC File , 44 KB - formative_v5i12e30762_app2.doc ]

Multimedia Appendix 3
Scale for daily urinary symptom scoring.
[DOC File , 35 KB - formative_v5i12e30762_app3.doc ]



XSL•FO
RenderX

## References


1. Wei JT, Calhoun E, Jacobsen SJ. Urologic diseases in america project: benign prostatic hyperplasia. J Urol 2008 May;179(5 Suppl):S75-S80. [doi: 10.1016/j.juro.2008.03.141] [Medline: 18405761]
2. Pesonen J, Vernooij R, Cartwright R, Aoki Y, Agarwal A, Mangera A, et al. The impact of nocturia on falls and fractures: a systematic review and meta-analysis. J Urol 2020 Apr;203(4):674-683. [doi: 10.1097/JU.0000000000000459] [Medline: 31347956]
3. Coyne K, Wein A, Tubaro A, Sexton C, Thompson C, Kopp Z, et al. The burden of lower urinary tract symptoms: evaluating the effect of LUTS on health-related quality of life, anxiety and depression: EpiLUTS. BJU Int 2009 Apr;103 Suppl 3:4-11. [doi: 10.1111/j.1464-410X.2009.08371.x] [Medline: 19302497]
4. Lerner LB, McVary KT, Barry MJ, Bixler BR, Dahm P, Das AK, et al. Management of lower urinary tract symptoms attributed to benign prostatic hyperplasia: AUA GUIDELINE PART I-initial work-up and medical management. J Urol 2021 Oct;206(4):806-817. [doi: 10.1097/JU.0000000000002183] [Medline: 34384237]
5. Welk B, McArthur E, Fraser L, Hayward J, Dixon S, Hwang YJ, et al. The risk of fall and fracture with the initiation of a prostate-selective α antagonist: a population based cohort study. BMJ 2015 Oct 26;351:h5398 [FREE Full text] [doi: 10.1136/bmj.h5398] [Medline: 26502947]
6. Cho ST, Kim JS, Noh J, Moon HS, Min SK, Bae S, et al. Characteristics of inappropriate multiple medication use in older urological outpatients. Arch Gerontol Geriatr 2019;83:61-65. [doi: 10.1016/j.archger.2019.03.024] [Medline: 30953962]
7. Yuan J, Mao C, Wong S, Yang Z, Fu X, Dai X, et al. Comparative effectiveness and safety of monodrug therapies for lower urinary tract symptoms associated with benign prostatic hyperplasia: a network meta-analysis. Medicine (Baltimore) 2015 Jul;94(27):e974 [FREE Full text] [doi: 10.1097/MD.0000000000000974] [Medline: 26166130]
8. Barry MJ, Williford WO, Chang Y, Machi M, Jones KM, Walker-Corkery E, et al. Benign prostatic hyperplasia specific health status measures in clinical research: how much change in the American Urological Association symptom index and the benign prostatic hyperplasia impact index is perceptible to patients? J Urol 1995 Nov;154(5):1770-1774. [doi: 10.1016/s0022-5347(01)66780-6] [Medline: 7563343]
9. Kobayashi M, Tokue A, Morita T. Discontinuation of tamsulosin treatment in men with lower urinary tract symptoms: a pilot study. Urol Int 2006;76(4):304-308. [doi: 10.1159/000092052] [Medline: 16679830]
10. Yokoyama T, Watanabe T, Saika T, Nasu Y, Kumon H, Miyaji Y, et al. Natural course of lower urinary tract symptoms following discontinuation of alpha-1-adrenergic blockers in patients with benign prostatic hyperplasia. Int J Urol 2007 Jul;14(7):598-601. [doi: 10.1111/j.1442-2042.2007.01784.x] [Medline: 17645601]
11. Chung JH, Lee JY, Kang DH, Jo JK, Lee JW, Lee SH, et al. Evaluation of patient outcome after discontinuation of alfuzosin treatment for benign prostatic hyperplasia: a multicentre, prospective study. Int J Clin Pract 2013 Sep;67(9):870-875. [doi: 10.1111/ijcp.12108] [Medline: 23952466]
12. Son H, Cho S, Park S, Kang J, Kim C, Kim H. A retrospective study of clinical outcomes of α-blocker or finasteride monotherapy followed by combination therapy: determination of the period of combination therapy of α-blocker and finasteride. Int J Clin Pract 2013 Apr;67(4):351-355. [doi: 10.1111/ijcp.12045] [Medline: 23521327]
13. Lee J, Kang D, Park S, Lee S, Kim Y, Choi H, et al. Effect of discontinuation of tamsulosin in korean men with benign prostatic hyperplasia taking tamsulosin and dutasteride: an open-label, prospective, randomized pilot study. Low Urin Tract Symptoms 2012 Jan;4(1):35-40. [doi: 10.1111/j.1757-5672.2011.00109.x] [Medline: 26676457]
14. Nickel JC, Barkin J, Koch C, Dupont C, Elhilali M. Finasteride monotherapy maintains stable lower urinary tract symptoms in men with benign prostatic hyperplasia following cessation of alpha blockers. Can Urol Assoc J 2008 Feb;2(1):16-21 [FREE Full text] [doi: 10.5489/cuaj.520] [Medline: 18542722]
15. Lin VC, Liao C, Kuo H. Progression of lower urinary tract symptoms after discontinuation of 1 medication from 2-year combined alpha-blocker and 5-alpha-reductase inhibitor therapy for benign prostatic hyperplasia in men--a randomized multicenter study. Urology 2014 Feb;83(2):416-421. [doi: 10.1016/j.urology.2013.09.036] [Medline: 24332123]
16. Nichol MB, Knight TK, Wu J, Barron R, Penson DF. Evaluating use patterns of and adherence to medications for benign prostatic hyperplasia. J Urol 2009 May;181(5):2214-21; discussion 2221. [doi: 10.1016/j.juro.2009.01.033] [Medline: 19296986]
17. Erickson BA, Lu X, Vaughan-Sarrazin M, Kreder KJ, Breyer BN, Cram P. Initial treatment of men with newly diagnosed lower urinary tract dysfunction in the Veterans Health Administration. Urology 2014 Feb;83(2):304-309 [FREE Full text] [doi: 10.1016/j.urology.2013.09.042] [Medline: 24286603]
18. Oelke M, Becher K, Castro-Diaz D, Chartier-Kastler E, Kirby M, Wagg A, et al. Appropriateness of oral drugs for long-term treatment of lower urinary tract symptoms in older persons: results of a systematic literature review and international consensus validation process (LUTS-FORTA 2014). Age Ageing 2015 Sep;44(5):745-755 [FREE Full text] [doi: 10.1093/ageing/afv077] [Medline: 26104505]
19. O'Mahony D, O'Sullivan D, Byrne S, O'Connor MN, Ryan C, Gallagher P. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. Age Ageing 2015 Mar;44(2):213-218 [FREE Full text] [doi: 10.1093/ageing/afu145] [Medline: 25324330]
20. Pereira Azevedo N, Gravas S, de la Rosette J. Mobile health in urology: the good, the bad and the ugly. J Clin Med 2020 Apr 03;9(4):1016 [FREE Full text] [doi: 10.3390/jcm9041016] [Medline: 32260206]




XSL•FO
RenderX

21. Rathbone A, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
22. Hantsoo L, Criniti S, Khan A, Moseley M, Kincler N, Faherty LJ, et al. A mobile application for monitoring and management of depressed mood in a vulnerable pregnant population. Psychiatr Serv 2018 Jan 01;69(1):104-107 [FREE Full text] [doi: 10.1176/appi.ps.201600582] [Medline: 29032705]
23. Portz JD, Vehovec A, Dolansky MA, Levin JB, Bull S, Boxer R. The development and acceptability of a mobile application for tracking symptoms of heart failure among older adults. Telemed J E Health 2018 Feb;24(2):161-165 [FREE Full text] [doi: 10.1089/tmj.2017.0036] [Medline: 28696832]
24. Bauer S, Breyer B, Oni-Orisan A, Steinman M, Sim I, McCulloch C, et al. PERSONAL: feasibility study protocol for placebo-controlled, randomized n-of-1 trials of tamsulosin for lower urinary tract symptoms. Front Digit Health 2020 Jun 26;2:7 [FREE Full text] [doi: 10.3389/fdgth.2020.00007] [Medline: 34713020]
25. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
26. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
27. Creswell J, Plano CV. Designing and Conducting Mixed Methods Research. Thousand Oaks: SAGE Publications Inc; 2018.
28. Kravitz R, Schmid C, Marois M, Wilsey B, Ward D, Hays R, et al. Effect of mobile device-supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. JAMA Intern Med 2018 Oct 01;178(10):1368-1377 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3981] [Medline: 30193253]
29. Moore CG, Carter RE, Nietert PJ, Stewart PW. Recommendations for planning pilot studies in clinical and translational research. Clin Transl Sci 2011 Oct;4(5):332-337 [FREE Full text] [doi: 10.1111/j.1752-8062.2011.00347.x] [Medline: 22029804]
30. Viechtbauer W, Smits L, Kotz D, Budé L, Spigt M, Serroyen J, et al. A simple formula for the calculation of sample size in pilot studies. J Clin Epidemiol 2015 Nov;68(11):1375-1379. [doi: 10.1016/j.jclinepi.2015.04.014] [Medline: 26146089]
31. Giuse N, Koonce T, Kusnoor SV, Prather AA, Gottlieb LM, Huang LC, et al. Institute of medicine measures of social and behavioral determinants of health: a feasibility study. Am J Prev Med 2017 Feb;52(2):199-206 [FREE Full text] [doi: 10.1016/j.amepre.2016.07.033] [Medline: 27659121]
32. Reeve E, Low L, Shakib S, Hilmer S. Development and validation of the revised Patients' Attitudes Towards Deprescribing (rPATD) questionnaire: versions for older adults and caregivers. Drugs Aging 2016 Dec;33(12):913-928. [doi: 10.1007/s40266-016-0410-1] [Medline: 27785734]
33. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida: CRC Press; 1996.
34. Weinfurt K, Griffith J, Flynn K, Cella D, Bavendam T, Wiseman J, LURN Study Group. The comprehensive assessment of self-reported urinary symptoms: a new tool for research on subtypes of patients with lower urinary tract symptoms. J Urol 2019 Jun;201(6):1177-1183 [FREE Full text] [doi: 10.1097/JU.0000000000000140] [Medline: 30730410]
35. Hays RD, Spritzer KL, Schalet BD, Cella D. PROMIS-29 v2.0 profile physical and mental health summary scores. Qual Life Res 2018 Jul;27(7):1885-1891 [FREE Full text] [doi: 10.1007/s11136-018-1842-3] [Medline: 29569016]
36. Voils CI, Hoyle RH, Thorpe CT, Maciejewski ML, Yancy WS. Improving the measurement of self-reported medication nonadherence. J Clin Epidemiol 2011 Mar;64(3):250-254 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.07.014] [Medline: 21194887]
37. Adam A, Jain A, Pletnikova A, Bagga R, Vita A, N Richey L, et al. Use of a mobile app to augment psychotherapy in a community psychiatric clinic: feasibility and fidelity trial. JMIR Form Res 2020 Jul 03;4(7):e17722 [FREE Full text] [doi: 10.2196/17722] [Medline: 32618572]
38. Nápoles AM, Santoyo-Olsson J, Chacón L, Stewart AL, Dixit N, Ortiz C. Feasibility of a mobile phone app and telephone coaching survivorship care planning program among Spanish-speaking breast cancer survivors. JMIR Cancer 2019 Jul 09;5(2):e13543 [FREE Full text] [doi: 10.2196/13543] [Medline: 31290395]
39. Lepor H. Pathophysiology of lower urinary tract symptoms in the aging male population. Rev Urol 2005;7 Suppl 7:S3-11 [FREE Full text] [Medline: 16986059]
40. Agarwal A, Eryuzlu LN, Cartwright R, Thorlund K, Tammela TL, Guyatt GH, et al. What is the most bothersome lower urinary tract symptom? Individual- and population-level perspectives for both men and women. Eur Urol 2014 Jun;65(6):1211-1217 [FREE Full text] [doi: 10.1016/j.eururo.2014.01.019] [Medline: 24486308]
41. Zhang L, Zhu L, Xu T, Lang J, Li Z, Gong J, et al. A population-based survey of the prevalence, potential risk factors, and symptom-specific bother of lower urinary tract symptoms in adult Chinese women. Eur Urol 2015 Jul;68(1):97-112. [doi: 10.1016/j.eururo.2014.12.012] [Medline: 25572826]
42. Everaert K, Anderson P, Wood R, Andersson F, Holm-Larsen T. Nocturia is more bothersome than daytime LUTS: results from an observational, real-life practice database including 8659 European and American LUTS patients. Int J Clin Pract 2018 Jun;72(6):e13091. [doi: 10.1111/ijcp.13091] [Medline: 29767479]
43. Egan K. The epidemiology of benign prostatic hyperplasia associated with lower urinary tract symptoms: prevalence and incident rates. Urol Clin North Am 2016 Aug;43(3):289-297. [doi: 10.1016/j.ucl.2016.04.001] [Medline: 27476122]




XSL•FO
RenderX

44. Lillie EO, Patay B, Diamant J, Issell B, Topol EJ, Schork NJ. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? Per Med 2011 Mar;8(2):161-173 [FREE Full text] [doi: 10.2217/pme.11.7] [Medline: 21695041]
45. Davidson K, Cheung Y, McGinn T, Wang Y. Expanding the role of N-of-1 trials in the precision medicine era: action priorities and practical considerations. NAM Perspectives. 2018. URL: https://nam.edu/expanding-the-role-of-n-of-1-trials-in-the-precision-medicine-era-action-priorities-and-practical-considerations/ [accessed 2021-11-12]
46. Wayne N, Ritvo P. Smartphone-enabled health coach intervention for people with diabetes from a modest socioeconomic strata community: single-arm longitudinal feasibility study. J Med Internet Res 2014 Jun 06;16(6):e149 [FREE Full text] [doi: 10.2196/jmir.3180] [Medline: 24907918]
47. Edelman M, Jellema P, Hak E, Denig P, Blanker MH. Patients' attitudes towards deprescribing alpha-blockers and their willingness to participate in a discontinuation trial. Drugs Aging 2019 Dec;36(12):1133-1139 [FREE Full text] [doi: 10.1007/s40266-019-00712-6] [Medline: 31515710]

## Abbreviations

**LUTS:** lower urinary tract symptoms
**mHealth:** mobile health
**PERSONAL:** Placebo–Controlled, Randomized, Patient-Selected Outcomes, N-of-1 Trials
**PROMIS:** Patient-Reported Outcomes Measurement Information System
**REDCap:** Research Electronic Data Capture
**rPATD:** revised Patients' Attitudes Toward Deprescribing

*Edited by G Eysenbach; submitted 28.05.21; peer-reviewed by M Blanker, B Nievas Soriano; comments to author 30.07.21; revised version received 08.10.21; accepted 13.10.21; published 10.12.21.*

*Please cite as:*
*Lee AW, Kenfield SA, Wang EY, Enriquez A, Oni-Orisan A, Steinman MA, Sim I, Breyer BN, Bauer SR*
*Tracking Lower Urinary Tract Symptoms and Tamsulosin Side Effects Among Older Men Using a Mobile App (PERSONAL): Feasibility and Usability Study*
*JMIR Form Res 2021;5(12):e30762*
*URL: https://formative.jmir.org/2021/12/e30762*
*doi:10.2196/30762*
*PMID:34889745*

©Austin W Lee, Stacey A Kenfield, Elizabeth Y Wang, Anthony Enriquez, Akinyemi Oni-Orisan, Michael A Steinman, Ida Sim, Benjamin N Breyer, Scott R Bauer. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# An Alternative to the Light Touch Digital Health Remote Study: The Stress and Recovery in Frontline COVID-19 Health Care Workers Study

Sarah M Goodday[1,2], PhD; Emma Karlin[1], MPH; Alexandria Alfarano[3], MA; Alexa Brooks[1], MSc, RD; Carol Chapman[1], MPH; Rachelle Desille[1], BA; Daniel R Karlin[1,4,5], MA, MD; Hoora Emami[6], BA; Nancy Fugate Woods[7], BSN, PhD, FAAN; Adrien Boch[8], MSc; Luca Foschini[8], PhD; Mackenzie Wildman[8], PhD; Francesca Cormack[9,10], PhD; Nick Taptiklis[9], BA; Abhishek Pratap[11,12,13,14], PhD; Marzyeh Ghassemi[12,15,16], PhD; Anna Goldenberg[12,17,18,19], PhD; Sujay Nagaraj[12,17], BSc; Elaine Walsh[7], PhD, RN, PMHCNS-BC; Stress And Recovery Participants[1]; Stephen Friend[1,2], MD, PhD

[1]4YouandMe, Seattle, WA, United States
[2]Department of Psychiatry, University of Oxford, Oxford, United Kingdom
[3]Nationwide Children's Hospital, Columbus, OH, United States
[4]MindMed, New York, NY, United States
[5]Tufts University School of Medicine, Boston, MA, United States
[6]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[7]School of Nursing, University of Washington, Seattle, WA, United States
[8]Evidation Health Inc, San Mateo, CA, United States
[9]Cambridge Cognition, Cambridge, United Kingdom
[10]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom
[11]Krembil Center for Neuroinformatics, Center for Addiction and Mental Health, Toronto, ON, Canada
[12]Vector Institute, Toronto, ON, Canada
[13]University of Washington, Seattle, WA, United States
[14]King's College London, London, United Kingdom
[15]Institute for Medical Engineering and Science, MIT, Cambridge, MA, United States
[16]Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, United States
[17]The Hospital for Sick Children, Toronto, ON, Canada
[18]Department of Computer Science, University of Toronto, Toronto, ON, Canada
[19]Canadian Institute for Advanced Research, Toronto, ON, Canada

**Corresponding Author:**
Sarah M Goodday, PhD
4YouandMe
2901 Third Ave Suite 330
Seattle, WA, 98121
United States
Phone: 1 2069288254
Email: sarah@4youandme.org

## Abstract

**Background:** Several app-based studies share similar characteristics of a light touch approach that recruit, enroll, and onboard via a smartphone app and attempt to minimize burden through low-friction active study tasks while emphasizing the collection of passive data with minimal human contact. However, engagement is a common challenge across these studies, reporting low retention and adherence.

**Objective:** This study aims to describe an alternative to a light touch digital health study that involved a participant-centric design including high friction app-based assessments, semicontinuous passive data from wearable sensors, and a digital engagement strategy centered on providing knowledge and support to participants.



XSL•FO
RenderX

**Methods:** The Stress and Recovery in Frontline COVID-19 Health Care Workers Study included US frontline health care workers followed between May and November 2020. The study comprised 3 main components: (1) active and passive assessments of stress and symptoms from a smartphone app, (2) objective measured assessments of acute stress from wearable sensors, and (3) a participant codriven engagement strategy that centered on providing knowledge and support to participants. The daily participant time commitment was an average of 10 to 15 minutes. Retention and adherence are described both quantitatively and qualitatively.

**Results:** A total of 365 participants enrolled and started the study, and 81.0% (n=297) of them completed the study for a total study duration of 4 months. Average wearable sensor use was 90.6% days of total study duration. App-based daily, weekly, and every other week surveys were completed on average 69.18%, 68.37%, and 72.86% of the time, respectively.

**Conclusions:** This study found evidence for the feasibility and acceptability of a participant-centric digital health study approach that involved building trust with participants and providing support through regular phone check-ins. In addition to high retention and adherence, the collection of large volumes of objective measured data alongside contextual self-reported subjective data was able to be collected, which is often missing from light touch digital health studies.

**Trial Registration:** ClinicalTrials.gov NCT04713111; https://clinicaltrials.gov/ct2/show/NCT04713111

***(JMIR Form Res 2021;5(12):e32165)*** doi:10.2196/32165

**KEYWORDS**

stress; wearable; digital health; frontline; COVID-19; health care worker; alternative; design; app; assessment; sensor; engagement; support; knowledge

## Introduction

### Background

The ubiquity of smartphones and the growing availability of wearable sensors has enabled a new era of digital health research [1-3]. The importance of this new form of remote health research has never been more apparent in light of the COVID-19 pandemic, which posed unprecedented challenges for the conduct of *traditional research* involving face-to-face contact [4]. Wearable devices such as smartwatches, smart rings, and bands enable the passive collection of broad semicontinuous physiological and activity information. Smartphones enable the collection of passive information in the form of activity tracking, phone use, and social media patterns, and high-frequency active tasks, surveys, and ecological momentary assessments. The benefits of remote digital health research involving these digital technologies include access to large sample sizes, cost efficiency, elimination of the necessity for travel, and ease of data collection owing to the capabilities to collect passive data. Participants can be remotely enrolled via eConsent frameworks [5] and recruited by social media channels. Accordingly, recruitment, consent, and onboarding can be conducted entirely outside of the clinic or site through the convenience of smartphones. Further, remote digital technologies enable the collection of rich multimodal data involving self-reported subjective and objective measured indicators of disease at semicontinuous or high frequencies. Importantly, these high-resolution data captures are possible outside health care or research visits in real-world settings. Yet, maintaining participant engagement throughout a digital health longitudinal study has proven a challenge [6-8]. The rapport built and *safety net* provided by the in-person visit represent a difficult gap to fill, particularly with additional remote technology usability challenges.

Several large app-based studies are described in the literature [9-11]. These studies share similar characteristics of a *light touch* approach that recruit, enroll, and onboard via a smartphone app and attempt to minimize burden through low-friction active study tasks while emphasizing the collection of passive data. Concerningly, there is strikingly low retention and adherence rates in app-based remote studies. Over half of participants tend to drop out after the first week of participation, while attrition and adherence differs significantly by important sociodemographic factors [6]. In addition to problems with engagement, there are common selection biases across studies tending to enroll White, university/college-educated participants with higher rates of women, reflecting nongeneralizable samples [6]. Further, patients more likely to use digital health trackers are more adherent to chronic disease medication use, suggesting those unlikely to engage in some digital tools may reflect less healthy populations [11]. Predictors of low digital engagement include lack of usability and accessibility, participant privacy and security concerns, perceived utility and motivation, and lack of support [7,8,12]. Although the *light touch* approach minimizes human contact with participants through fully remote enrollment and follow-up via an app, the lack of "human-in-the-loop" and clear value proposition for participants may inadvertently lower their engagement. Digital health cohorts being recruited via a clinic referral compared to recruitment conducted entirely through an app show higher rates of retention and adherence [6]. The *light touch* approach also minimizes the collection of self-reported subjective data to reduce participant burden and in turn attrition, yet this information is crucial to validate objective measured information. This is particularly important given that the field is in an early phase with the need to validate objective measured sensor readouts.

The use of participant incentives and rewards still prevail as one of the most used components for a successful engagement strategy. The use of smartphones makes personalized rewards and reminders possible [13]. Leveraging behavioral psychology, informed strategies for reward scheduling, smartphones can incentivize adherence through rewards for study task completion. Participant tailored push notifications for task



XSL•FO
RenderX

reminders can be implemented and have been noted as preferred by participants in digital health research [14]. Beyond incentives, treating participants differently than the traditional blinded *participant* and including them as codrivers in the research process, could prove a powerful way to engage, retain, and accelerate learning for long-term engagement.

The shift to participatory research that tends to involve patient advisory groups who provide input on study design documents such as consent forms [15] and study protocols is already being conducted. However, in the context of digital health and in the use of digital technologies, these participatory models tend to only include users after the relevant technology has been developed. These approaches can be extended further, described by some as "user centered designs," where participants and patients might be included from the early design to implementation phases and might help shape how the technology is used [16]. Further, participant-centered initiatives [17,18] aim to include participants as equal partners in the entire research process in testing the feasibility of digital technologies for health and wellness. Two recent app-based studies [14] involving a patient- and citizen-centric design, achieved increased retention and adherence when compared to that typically reported in such studies, which demonstrates the promise of these more patient-centered approaches.

Here, we describe an example of an alternative to the *light touch* digital health study that involves active app- and wearable-based assessments coupled with an extensive digital engagement participant-centric strategy. Our objective was to demonstrate that participant-centric engagement approaches might enable a digital health study with improved participant experience, likelihood to retain in the study, and adherence to protocols.

### A Case Example: The Stress and Recovery in Frontline COVID-19 Health Care Workers Study

The COVID-19 pandemic has caused unprecedented stress on health care systems in affected countries and, in particular, on the health care workers working directly with patients with COVID-19. Health care providers faced and continue to face numerous stressors relating to higher risk of COVID-19 exposure, unpredictable work shifts and shifting health care policies, and worry over family member risk. This frontline health care population provides a unique example to test the feasibility of detecting stress using wearable-based technology and smartphone apps. Further, this population could inform understanding of how stress impacts susceptibility to infection, given the damaging impact of stress on our immune system [19]. The accurate measurement of stress responses in real time and in naturalistic settings has so far been a challenge [20], limiting our understanding of how different facets of acute or sustained stress increases susceptibility to breakdown and disease. Studies of stress in frontline populations exist [21], but focus on self-reported stress during aperiodic frequencies, as opposed to high-resolution approaches using digital technologies. Further, studies aimed at objective stress detection using sensor-based tools, irrespective of population, are typically limited to controlled experimental settings (eg, [22]). There are few studies applying wearable technologies used in real-world settings for the detection of stress. Among the few pilot studies that have been conducted [20,23-25], wearable technologies are showing promise for detecting shifts in health status, stress, and well-being across different populations.

The COVID-19 pandemic reflects a unique natural experimental condition where frontline workers were exposed to substantial stress beyond that already present in their pre–COVID-19 day-to-day work environment. Their on-shift time provided a naturally occurring "stress on" period, while their off-shift time provided a "stress off" period and an opportunity to follow an individual's recovery from stress. The aims of the Stress and Recovery Study were to:

- Assess the feasibility of a participant-centric digital approach to collect both participant-reported subjective and objective measured longitudinal high-resolution data on immediate stress responses, intermediate signs of stress, recovery from stress, and COVID-19 infectivity by engaging frontline health care workers in the use of digital sensors
- Determine the feasibility of detecting and tracking changes in immediate, intermediate stress, and recovery from stress in frontline health care workers working with patients in the COVID-19 pandemic environment

## Methods

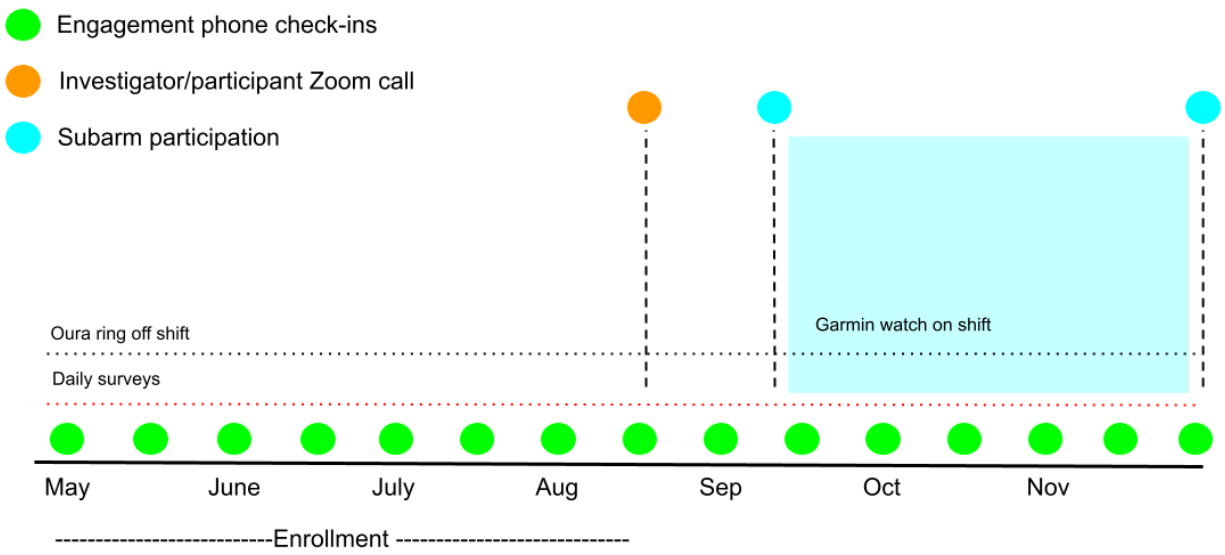
### Overview

This study involved engaging US frontline health care workers from a variety of locations who were either working directly with patients with COVID-19 or whose work routines were shifted as a result of the COVID-19 pandemic. This study used a participant-centric design [16-18] (Figure 1) that comprised three primary components and a follow-up period of 4 to 6 months: (1) patient-reported active and passive assessments of stress and symptoms from a smartphone app; (2) objective measured assessments of acute stress from wearable sensors; and (3) a participant codriven engagement strategy that centered on providing knowledge and support involving regular check-in phone calls with study staff, using participant feedback in real time to improve and fine-tune the research protocol to improve participant experience, and the addition of new study subarms. A virtual event was held where researchers and participants who wished to join were able to engage in an online group setting (anonymously) halfway through enrollment. This study was approved by the Institutional Review Board, Advarra (4UCOVID1901, Pro00043205) and was registered with ClinicalTrials.Gov (NCT04713111).



XSL•FO
RenderX

**Figure 1.** Stress and Recovery study design.




## Recruitment and Onboarding

Frontline health care workers were recruited from May to August 2020. A multipronged recruitment approach was developed that involved engaging trusted leaders from organizations and sites with high outreach to our target population (eg, American Association of Critical Care Nurses), engaging supervisors at selected health care institutions, and a social media campaign (eg, Facebook, Twitter, and Linkedin). Tailored workplace-specific recruitment materials were developed and, during the enrollment period, assessed in real time how participants were learning about the existence of the study to understand the most effective recruitment strategies. Recruitment materials were distributed through workplace-specific newsletters and websites, and through our social media channels.

## Population

Frontline health care workers were invited to participate including medical doctors, doctors of osteopathy, physician assistants, registered nurses, advanced practice registered nurses, and other allied health care workers. Inclusion criteria were must be working directly with patients with COVID-19, slated to do so within the next 2 weeks, or work routines have been moderately or extremely impacted by the COVID-19 pandemic; older than 18 years; able to speak, write, and read English; able to provide informed consent; no known SARS-CoV-2 current or past infection; and must own a personal iOS mobile phone (OS11 and above) with willingness to download and use the study apps and sync phone with all study sensors.

## App-Based Data Assessments

Active study data were collected and managed in a REDCap (Research Electronic Data Capture [26,27]) database, hosted and managed by the Center for International Emergency Medical Services [28]. REDCap is a secure, web-based software platform designed to support data capture for research studies. The MyCap app interface leverages REDCap and was used to produce a study app for the collection of participant-reported active data. Participants were instructed to download the MyCap app from the app store and register with the MyCap app through a unique QR Code provided to them by study staff.

Via the study app, participants were prompted to complete daily, weekly, every 2-week, monthly, and one time measures that involved sociodemographic factors, self-reported measures of daily perceived stress, intermediate signs of stress (sleep, mood, and cognition), additional health-related symptoms, influenza-like illness, individual characteristics, and cognitive active tasks (see Multimedia Appendix 1, Table S1 [29-51]). For the first half of the study ResearchKit's active tasks [29] (Trail Making, Reaction Time, and Spatial Span Memory) were used every day, rotating the tasks each day. For the second half of the study these were replaced with Cambridge Cognition's active tasks [30,31] (emotional bias test, psychomotor vigilance, and n-back test) rotating every other day. The study intended to include the Cambridge Cognition active tasks from the outset; however, their implementation was delayed while establishing the technical integration with the MyCap app.

The study sample was initially limited to iOS users because of anticipated nonfunctionality with the ResearchKit apps but intended to include Android users when shifting to the Cambridge Cognition tasks. However, upon pilot-testing the REDCap app with the first set of enrolled Android users, it was found that the app itself had compatibility issues. Therefore, a subset of 12 Android users enrolled in the study who did not participate in the app-based assessments.

The daily burden for completing app-based assessments was estimated at an average of 5 minutes per day (minimum daily burden 3.5 minutes, maximum daily burden 8.5 minutes), with some daily tasks taking longer and other days shorter, which depended on the cadence of the one time, weekly, every 2 week, or monthly measures. This estimate was calculated from the expected task length, not from timed participant data. A schedule was produced that spread out the one-time measures on different



XSL•FO
RenderX

days within the first month, while weekly and every 2-week tasks were scheduled on different days to balance the daily burden.

Participants were given the option to download two third-party apps as part of their participation: HealthMode Cough app and RescueTime. The cough app was used to capture momentary cough during study follow-up and RescueTime as a measure of screen time (eg, time spent in and category of apps) as a proxy for objective stress and mood.

### Wearable Assessments

Participants were mailed an Oura smart ring. The Oura Ring 2 [52] is made of a light durable titanium shell and includes a temperature sensor, a gyroscope, a 3D accelerometer, and an infrared optical pulse sensor. There was a one-time setup process where the participant synced their ring to the Oura smartphone app that they were instructed to download. Throughout the study, the participant was instructed to open the Oura smartphone app to sync data off the ring over Bluetooth. The sensor collects a variety of nighttime data streams such as heart rate, heart rate variability, and objective sleep quality measures (Multimedia Appendix 1, Table S1). Participants were provided with their own symptom summaries via the Oura app. Participants were instructed to wear the ring only while off shift owing to potential infectivity risks generally associated with the ring wearing while at work in health care settings, which was especially relevant as participants were actively working with patients. This did not meaningfully hamper relevant data collection because we were most interested in measuring parameters associated with recovery from stress while participants were off shift. It was expected that the daily burden associated with using sensors, remembering to charge them, and working through sensor issues to be approximately 2 to 4 minutes per day.

### Engagement

The study engagement strategy centered on providing information and support to participants while engaging them as codrivers of the research. This strategy had two aims: (1) to engage participants in the use of the study digital devices for optimal adherence and (2) to engage participants in the design of the study. Information was provided to participants through enabling insights into measurements of health through the study sensor apps and discussing this with participants in terms of how to interpret this information and what could be learned from it. Support was provided in a variety of channels through the biweekly check-ins, listening to participant feedback and making real-time protocol changes, and providing a variety of tools or resources in an attempt to give back (online resources for stress management) and stress-reducing tools. For example, participants were offered a YELL-IT tool where they could call a number and leave an anonymous voice message of their choice that could include any release that might offer benefit (ranting, yelling, journaling, etc). The records of these phone calls were immediately programmatically deleted, and the voice messages themselves were not recorded.

### Check-in Calls

Research staff labelled as *engagement specialists* contacted study participants every week for the first month of study participation and every 2 weeks thereafter until study completion. Engagement specialists had clinical research backgrounds and experience with working with participants. These study staff were trained in the use of digital technologies and served the role to support and engage participants in their digital experience. The check-in calls with engagement specialists served three purposes: (1) to support participants in their study participation, troubleshoot technical problems, and build rapport; (2) to discuss, understand, and collect information on study experience; and (3) to discuss, understand, and collect information on study exposure and outcome information, which in this context was the experience of stress from working on the frontline in the COVID-19 pandemic environment. The check-in calls served as a venue to gather deep insights about the participant experience in general, specifically around interacting with digital sensors for stress and health tracking. Engagement specialists reviewed adherence data prior to check-ins to probe participant-specific study challenges. Check-in calls were expected to range in time depending on the need of the participant, but the initially allotted time was up to 60 minutes for each of these calls (see Results section).

Engagement specialists conducted exit interviews by phone at the end of the study; interviews lasted approximately 1 hour. Participants were asked open-ended questions relating to their work in the pandemic environment and about features of the study that might help them and others in the future.

### Addition of New Optional Subarms

Halfway through enrollment participants were invited to participate in 1 to 3 new study subarms. The subarms were announced during the joint investigator and participant Zoom (Zoom Video Communications, Inc) meeting and during biweekly check-ins. These included an arm with a wearable smartwatch to be worn on shift, an arm with a lifestyle intervention, and an arm where hair cortisol was measured. Interested and eligible participants were sent a new REDCap link to consent to participate in these subarms.

For the wearable arm, participants were provided with Garmin Vivoactive 4 smartwatches [53] and were instructed to wear these continuously for 4 weeks (including while on shift) to capture on-shift objective measures of stress.

For the intervention arm, participants were able to self-select into a physical activity subarm or a meditation subarm for 4 weeks. Participants in the meditation arm were provided a free subscription to the Headspace app—a publicly available mindfulness app that offers guided meditation sessions among many other features aimed at improving mood, sleep, and stress. Participants were instructed to complete three to five or more mindfulness sessions per week. Participants in the physical activity subarm were provided with a resource comprising a variety of free online fitness classes and were instructed to engage in 30 minutes to 1 hour of physical exercise of their choice, 2 to 3 times a week or more.



XSL•FO
RenderX

For the hair cortisol arm, participants were sent hair sample collection kits with instructions to self-collect and send a hair sample back to the study team to provide a biological measure of chronic stress during the study period [54]. Cortisol concentrations were extracted using a standard kit (ie, ELISA) by the laboratory services at the School of Nursing at the University of Washington (please see Multimedia Appendix 1, Methods for a detailed description of the subarms).

### Joint Participant-Investigator Video Meeting

A joint participant-investigator Zoom call meeting halfway through enrollment was held. The purpose of this meeting was to give participants a chance to meet the study team in person (virtually), ask questions, and give feedback. The study team gave study updates on progress and introduced the new optional study arms. Participants' confidentiality was maintained by using anonymous mode features of the video call platform. Although participants were anonymous in the call, they could all see and hear the study team, and could participate via the chat feature to ask questions and give feedback.

### Learning by Doing

The goal of the engagement approach was to change the participant experience from feeling like only a source of data, or a blinded "subject," to a supported project codriver and partner in the research. As this was a feasibility study, we engaged participant feedback from the study's start and implemented protocol changes during follow-up. Accordingly, participants directly helped shape the nature of how we asked app-based assessments, how we explained study-related details, and how we will design future remote digital health studies. Participants were invited to be coauthors of study-related published work.

### Total Participation Effort

It was estimated that the total effort for study participation was on average 10 to 15 minutes per day. Beyond app-based assessments, that on average took 5 minutes per day, additional activities included charging sensors, daily opening of the study apps and syncing of sensors, viewing the data from the associated apps, miscellaneous tech issues, check-in calls, and correspondence with an engagement specialist to schedule check-in calls and other study-related activities including exit interviews and the Zoom call. This time estimate was derived by study investigators and staff communication with participants on all daily activities as previously described.

### Compensation and Benefit

Participants were not offered any monetary incentives, nor were rewards in the form of points provided for study participation. However, participants were allowed to keep the Oura Ring and the Garmin smartwatch (in subarm participants) at the end of their participation.

### Analysis

Univariate descriptive analyses of cohort characteristics, retention, and adherence are reported. Survival probabilities using the Kaplan-Meier approach were calculated to display retention over the course of the study. Bivariate associations between cohort characteristics and adherence rates were conducted using chi-square, Fisher exact (for cell counts less than five), and *t* tests where appropriate. Mixed effects linear models were used to estimate changes in weekly mean adherence rates by group status using an autoregressive covariance matrix. Thematic study insights from qualitative data are described from participant and study staff feedback. Analyses were conducted using SPSS version 27 (IBM Corp) [55].

Retention was defined as completing the minimum follow-up, which was 4 months, or retained until the end of study, which was defined by a specific cutoff date. Adherence was defined as the number of tasks completed over the total number of available tasks that could be completed by participants' unique study time. For example, a participant with a total study time of 140 days (20 weeks) and having completed 100 daily surveys, 17 weekly surveys, and any Oura data upload (even if a partial day) for 130 days would have an adherence rate of 71.43% (100/140) for daily assessments, 85.00% (17/20) for weekly assessments, and 92.86% (130/140) for Oura wearable data, respectively.

## Results

### Description of the Cohort

The final study sample included 365 participants who enrolled in and started the study (Multimedia Appendix 1, Figure S1). The median age was 33 (range 20-67) years (Table 1). The majority of participants were female (n=325, 89.04%), White (n=302, 82.74%), and were registered nurses (n=325, 89.04%; Table 1). Participants were followed for a median follow-up of 112 (range 1-170) days during the period from May 1, 2020, to November 20, 2020. Primary reasons for exclusion were being an Android user, prior COVID-19 infection, and no direct patient care (Multimedia Appendix 1, Figure S1). Participants were located across 27 different US states, with the majority of participants working and residing (at the time of participation) in Washington (n=103, 34.68%), Minnesota (n=66, 22.22%), Massachusetts (n=38, 12.79%), Arizona (n=27, 9.09%), and Wisconsin (n=19, 6.34%). There were five reported cases of COVID-19 during study follow-up among the study participants.



XSL•FO
RenderX

**Table 1.** Characteristics of the cohort among those who enrolled, completed the study, and did not finish.

| | Enrolled and started the study (N=365), n (%) | Retained[a] (n=297), n (%) | DNF[b] (n=68)[c], n (%) | *P* value (retained vs DNF) |
|---|---|---|---|---|
| **Age (years)** | | | | .16[d] |
| 18-25 | 47 (12.88) | 37 (12.46) | 10 (14.71) | |
| 26-35 | 168 (46.03) | 129 (43.43) | 39 (57.35) | |
| 36-45 | 78 (21.37) | 67 (22.56) | 11 (16.18) | |
| >46 | 72 (19.73) | 64 (21.55) | 8 (11.76) | |
| **Gender** | | | | .85[d] |
| Female | 325 (89.04) | 264 (88.89) | 61 (89.71) | |
| Male | 40 (10.96) | 33 (11.11) | 7 (10.29) | |
| **Race** | | | | .35[e] |
| White | 302 (82.74) | 242 (81.48) | 60 (88.24) | |
| Black or African American | 8 (2.19) | 8 (2.69) | 0 (0.00) | |
| Asian/Pacific Islander | 27 (7.40) | 22 (7.41) | 5 (7.35) | |
| Native American or American Indian | 2 (0.55) | 1 (0.34) | 1 (1.47) | |
| More than one race | 21 (5.75) | 19 (6.40) | 2 (2.94) | |
| Unknown/not reported | 5 (1.37) | 5 (1.68) | 0 (0.00) | |
| **Ethnicity** | | | | .69[e] |
| Hispanic or Latino | 12 (3.29) | 9 (3.03) | 3 (4.41) | |
| Not Hispanic or Latino | 343 (93.97) | 279 (93.94) | 64 (94.12) | |
| Unknown/not reported | 10 (2.74) | 9 (3.03) | 1 (1.47) | |
| **Occupation** | | | | .39[e] |
| Registered nurse | 325 (89.04) | 264 (88.88) | 61 (89.71) | |
| Medical doctor | 5 (1.37) | 5 (1.68) | 0 (0.00) | |
| Medical assistant | 10 (2.74) | 7 (2.36) | 3 (4.41) | |
| Emergency medical services | 2 (0.55) | 1 (0.34) | 1 (1.47) | |
| Other[f] | 23 (10.96) | 20 (6.73) | 3 (4.41) | |

[a]Retained includes completing follow-up of 4 months or by study end cutoff date.

[b]DNF: did not finish.

[c]DNF includes participants lost to follow-up, dropped out, or withdrawn.

[d]Pearson chi-square tests.

[e]Fishers exact tests.

[f]Other occupations include social workers, respiratory therapist, surg/cardio tech, dentist, registered dietitian, or medical student.

## Recruitment

Upon screening, participants were asked “How did you hear about our study?” Of participants screened and enrolled into the study, over 50.00% (n=200) found out about the study through a recruitment email from their workplace department or floor, while approximately 40.00% (n=151) found out about the study through word of mouth. The remainder (<10.00%) found out about the study through national associations and workplace-specific newsletters and social media channels.

## Retention

Of the 365 enrolled participants, 81.37% (n=297) completed the study. Of the 68 participants who did not complete the study, 11.76% (n=8) were withdrawn (study team withdrew) due to no longer meeting inclusion criteria such as no longer working with patients, furloughed, or lost study sensors; 41.18% (n=28) dropped out (participant decided to end participation) due to reasons such as no longer interested, not enjoying the study or study sensor, or too much time commitment; and 51.47% (n=35) were lost to follow-up (could not be reached on recontact). The probability of retaining in the study for 1 month was 98.00%, while the probability of retaining in the study halfway through

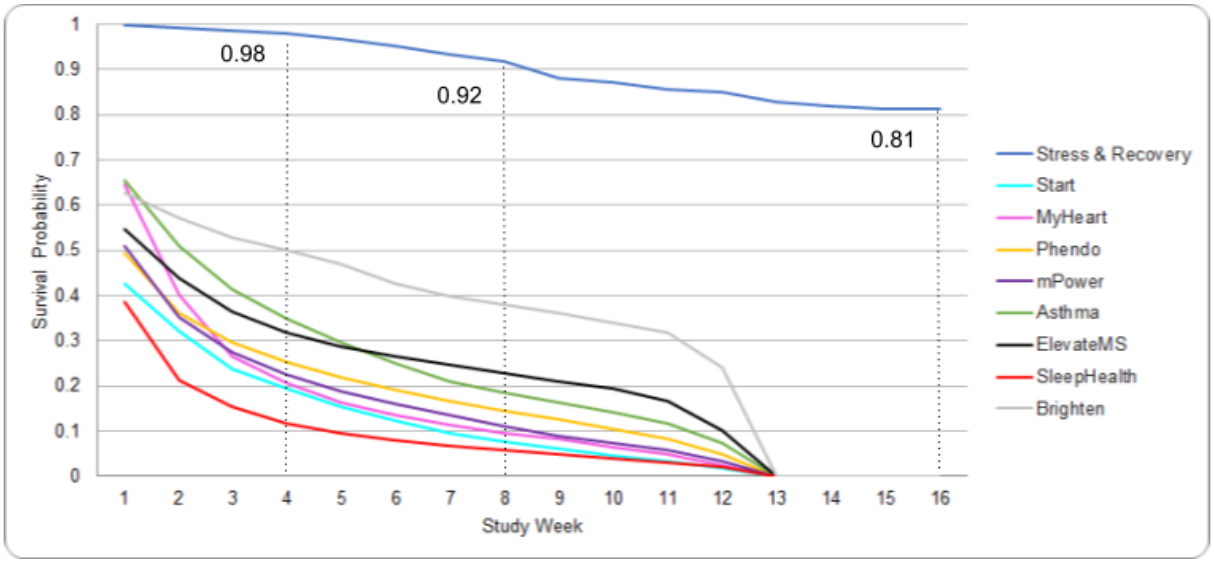


XSL•FO
RenderX

study time was 92.00% (Figure 2). Retention for the three study subarms can be found in Multimedia Appendix 1, Figure S2.

Sample characteristics were similar in participants who started the study compared to those retained in the study. There were higher proportions of younger individuals not completing the study compared to those retained, although this was not statistically significant (*P*=.16). Other sample characteristics were similar in participants who did not complete the study compared to those retained although cell counts were low across categories (Table 1).

**Figure 2.** Survival probability of retaining in the study. Additional data from Pratap et al [6]. Kaplan-Meier survival curves for the Stress and Recovery study and for 8 additional digital health app-based studies as described in Pratap et al [6]. Please interpret with caution. The survival probabilities from the 8 studies included in Pratap et al [6] included a mix of different study populations some including chronic disease populations and some healthy populations with different study durations.




## Adherence

Adherence calculations are presented for those participants who completed the study (Table 2). Although initially excluded, 12 Android users were enrolled when the study protocol switched to Cambridge Cognition tasks from ResearchKit's active tasks because of expected sensor compatibility. However, owing to troubleshooting problems with the MyCap app, these 12 individuals (minus 1 participant who did not complete the study) were unable to use the study app and were excluded from app-based adherence calculations.

Participants adhered to wearing the Oura smart ring and the Garmin smartwatch for an average of 90.60% and 90.42% of study time, respectively. App-based daily, weekly, every 2 weeks, and monthly surveys were completed on average 69.18%, 68.37% (range across different tasks 64.44%-71.86%), 72.86% (range across different tasks 72.42%-73.30%), and 68.82% (range across different tasks 68.05%-69.82%) of study time, respectively (Table 2). Every 2-week check-in phone calls were completed for an average of 75.62% of study time (Table 2 and Figure 3). The average check-in call length was approximately 14.5 minutes and ranged from 2 to 70 minutes. Measures scheduled on Fridays and Saturdays had consistently lower adherence compared to other days of the week. Average adherence was higher for the ResearchKit active tasks (80.59%) compared to the Cambridge Cognition tasks (56.49%). However, the ResearchKit tasks were integrated within the study app and were shorter in duration, while the Cambridge Cognition tasks had to be completed via an external web-based link, which may have contributed to lower adherence on these tasks. Weekly average adherence on daily app measures across age categories were similar, and not statistically significant ($F_3$=0.20; *P*=.89), although there was a trend where higher age categories demonstrated higher adherence on Oura Ring use ($F_3$=2.49; *P*=.06). Adherence across other sample characteristics was not explored owing to small sample sizes across categories.

Average adherence in the week after the joint participant-investigator Zoom call was held showed large increases for app-based daily surveys (82.93%), for Cambridge Cognition tasks (88.97%), and in Oura Ring use (97.89%).



XSL•FO
RenderX

**Table 2.** Adherence rates by study activity.

| | Full study period |
|---|---|
| Study app surveys, n | 286[a] |
| Daily survey[b], % | 69.18 |
| Weekly surveys, % | 68.37 |
| Biweekly surveys, % | 72.86 |
| Monthly surveys, % | 68.82 |
| ResearchKit tasks, n[c] | 164 |
| Cognitive active tasks, % | 80.59 |
| Cambridge cognition, n[d] | 289 |
| Cognition tasks, % | 56.49 |
| ES[e] check-ins, n | 297 |
| Biweekly check-ins[f], % | 75.62 |
| Oura Smart Ring, n[g] | 296 |
| Oura Ring use, % | 90.60 |
| On-shift wearable subarm, n | 95 |
| Garmin Smartwatch use, % | 90.42 |

[a]Excluding 11 participants who were Android users and unable to use the study app.

[b]All study app survey completion calculations exclude 12 participants (11 retained) with Android sensors who have no study app data

[c]ResearchKit active tasks were switched to Cambridge Cognition tasks on July 6, 2020; therefore, some participants did not receive at least 1 ResearchKit tasks as reflected by a smaller study sample size

[d]The higher sample size reflects a few Android users who were able to access the web-based Cambridge Cognition links.

[e]ES: engagement specialist.

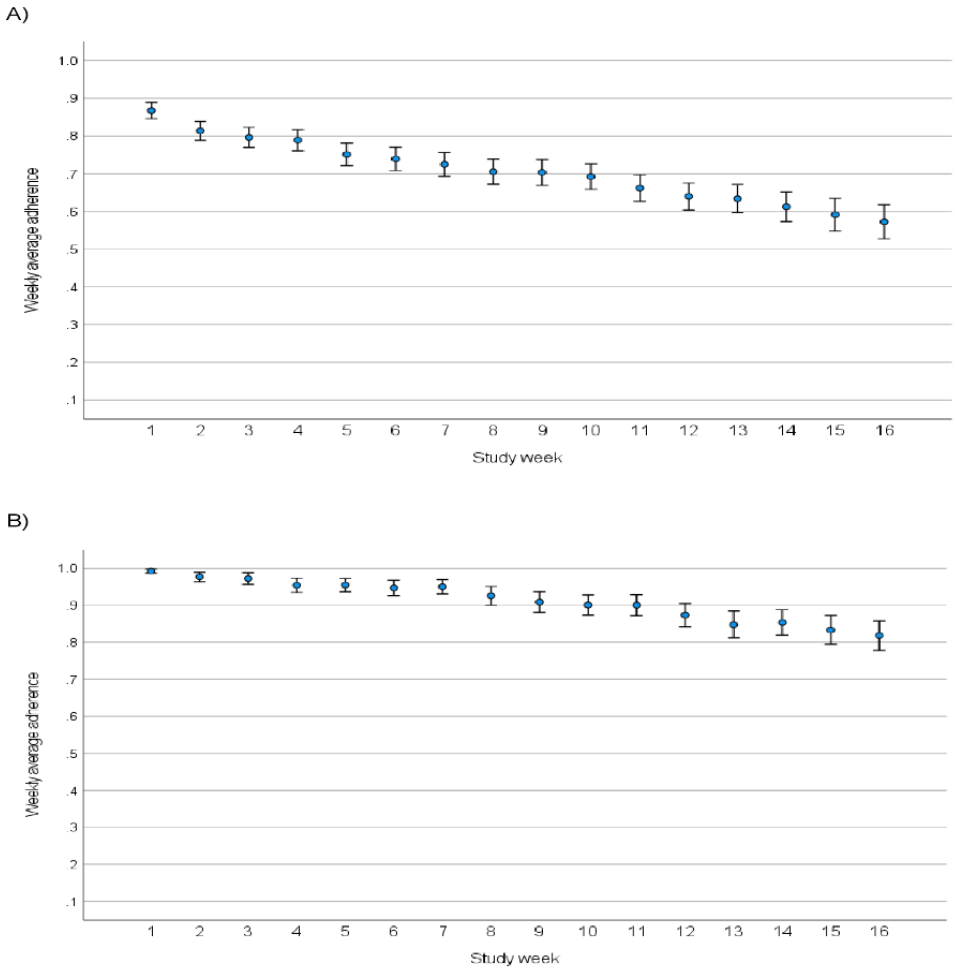
[f]Of the 297 retained participants, 2 completed 0 check-ins.

[g]One retained participant never synced their ring to the app.



XSL•FO
RenderX

**Figure 3.** Weekly average adherence and standard errors for daily app-based tasks (A) and Oura Ring use (B) in retained participants (n=297).




### Engagement Impact on Adherence

Although this study was not designated to explicitly test different aspects of the engagement approach on study acceptability and experience, to explore whether study-related events had an impact on adherence, weekly average adherence on active daily app-based surveys in participants enrolled prior to the joint participant-investigator Zoom call, and therefore had the opportunity to participate (n=246), and participants enrolled after the Zoom call, and therefore did not participate (n=39), were calculated. As seen in Multimedia Appendix 1, Figure S3, participants who were enrolled later in the enrollment period and did not have the opportunity to participate in the joint investigator-participant Zoom call showed lower weekly average adherence (mean 0.66, SD 0.22) on daily app-based tasks compared to those enrolled prior to this meeting (mean 0.71, SD 0.21); although, this was not statistically significant over study time ($F_1$=3.06; $P$=.81; Multimedia Appendix 1, Table S2).

### Learning by Doing: Insights Learned During Study Follow-up

Knowledge was gained from the check-in calls with engagement specialists and from the engagement specialists themselves through discussions with participants that fueled insights into study improvements. The insights learned can be found in Multimedia Appendix 1, Table S3. Protocol changes implemented can be found in Multimedia Appendix 1, Table S4. Key themes centered on were privacy (particularly on the surveillance nature of some app features (eg, RescueTime), usability, perceived utility, and knowledge of how to interpret sensor readouts, particularly the objective data (eg, heart rate and heart rate variability). Dedicating check-in calls to explain the purpose of individual measures and how data are used by third-party apps helped with these challenges. Although not an initial purpose of check-in calls, in discussing stress and symptom experience with participants, these calls also served as an outlet for some participants to discuss with someone outside their place of employment their COVID-19 frontline experience. Accordingly, these calls may have produced an inadvertent interventional effect.

## Discussion

### Main Findings

This study tested the feasibility of a participant-centered digital health study with a daily burden of 10 to 15 minutes in frontline health care workers. We found support for the feasibility and acceptability of this approach with 81% (n=297) retention, while average adherence for wearable sensor use and daily app-based assessments was approximately 90% and 70% of study time, respectively. This contrasts to typical reported retention and



XSL•FO
RenderX

adherence rates in digital medicine studies that tend to be lower than 50% [6] and that had much lower daily burdens, although the underlying populations of these studies are different, which makes direct comparisons difficult. In addition to high retention and adherence, the collection of large volumes of objective data alongside contextual self-reported subjective data was able to collect what is often missing from the *light touch* digital health study.

The COVID-19 pandemic has highlighted the increasing importance of being able to operate, communicate, and conduct research remotely and digitally. Yet, historically remote digital studies have been hampered by low retention and engagement of participants. This poses obvious challenges for the usability and generalizability of digital data and is also an early warning signal. Poor engagement at the research phase provides clues into the challenges we will face at the health care implementation phase. The engagement approach developed here involved three components: (1) supportive check-in calls with engagement specialists during follow-up, (2) a learning-by-doing approach that leveraged direct feedback from participants collected during check-ins to fuel real-time study improvements and participant experience, and (3) new study features and a virtual investigator-participant event. Findings from this feasibility study suggest these patient-centered strategies offer enough value to sufficiently engage participants without monetary/reward-based incentives. Although participants were offered to keep their Oura Ring and Garmin watch at the end of participation, no monetary incentives or point-based rewards systems were used. Both the increase in adherence after the joint participant-investigator Zoom call and higher daily adherence among those with the opportunity to participate in the Zoom call suggests that potential self-awareness and benefit from the sensors alone are not the only reason for sustained adherence to protocols. Further, this approach enabled the collection of rich objective measured data alongside participant self-reported subjective data. The common *light touch* digital health study often lacks adequate contextual self-reported subjective data to ground and validate the objectively collected information, which poses challenges in interpreting the data. These findings suggest that engaging participants in the appropriate way can enable a high burden study and the subsequent collection of needed contextual self-reported subjective data.

Digital sensors that can return symptoms back to the user enable a two-way learning experience in which participants learn about themselves in real time and provide insights to researchers, in contrast to traditional methodologies that collect data from participants in blind or shielded ways, where the data are then unveiled at the end of the study. The former enables accelerated learning at the pilot research phase: a learning-by-doing approach that leverages digital tools to enable participants to actively partake in and shape their digital research experience through tracking their own health data and discussing these data with study investigators in real time. However, the impact of tracking objective measures of health is largely unknown in terms of how being enabled to track personal objective data impacts the user. Further, wearable sensor companion apps have embedded nudges and prompts to shift behavior based on the collected data. Some participants noted frustration with both the returning of the Oura Ring collected symptom summaries and associated feedback prompts, as these were nonactionable, particularly in a health care professional population. On the contrary, others viewed these returned symptoms and nudges positively. Commonly, participants noted a desire to have more knowledge about interpreting the sensor readouts and were curious about other participants' data. Beyond digital literacy, there is complex knowledge and support required in the use of digital tools for stress and health tracking. The field is in an early phase of understanding how individuals from diverse populations will interact with digital technologies at home and in everyday life that will be needed for the successful implementation of digital approaches into health systems and for their use in transforming individualized care. Support in the use of these tools is a current gap. The notion of a *digital expert* or counselor, not dissimilar to genetic counselors and like the engagement specialists used here, may be one approach to bridge this gap both for digital research and digital health care.

## Limitations

The study population included predominantly White health care professionals who are nonrepresentative of non-White and non–health care provider populations in having higher than average knowledge of research and higher health literacy. It is unclear how the engagement approach might generalize in other populations. On the other hand, this population is a busy, high-stress population. In light of this, one could argue that this particular population would be difficult to engage in a high-friction study owing to work-life constraints. A nonprobability sample was included; therefore, selection biases may be present where participants enrolled may be more likely to engage or have interest in the use of wearable sensors for health tracking compared to those who were uninterested. A control group of participants who did not receive the adjacent engagement strategy were not included; therefore, we cannot imply causality of this approach or parse the different possible drivers of success on retention and adherence results. Further, the cost of the Oura Ring and Garmin Smartwatch is high (>US $500). Although no monetary incentives were offered, the opportunity to keep these devices could have impacted willingness to retain and adhere to protocols.

The participant-centered learning-by-doing approach used here worked well for a feasibility study where the primary purpose is to learn about how well a methodology or a tool will work for health-related purposes. However, these approaches may not be well suited to other types of studies that require more controlled data collection such as in randomized controlled trials. Enabling participants to track their own health outcomes while they are under study may increase risk of bias in the controlled study context through participants' awareness of symptoms, particularly through companion app nudges and prompts. Although, this hallmark challenge in traditional controlled studies largely reflects a risk to altering perceptions of the outcome of interest, which is how symptoms are traditionally measured. In the context of enabling individuals to be aware of their own measured objective signs of stress and disease, the importance of this traditional challenge becomes less clear. An additional possibility is that the implemented



XSL•FO
RenderX

engagement approach may have produced a positive interventional effect from either the returned symptoms or the every 2-week check-in calls, which could modulate the stress signal in the data.

Additional challenges of this approach encompass a time burden on researchers and staff. Both investigators and staff were highly engaged during the study follow-up. Each check-in call took on average 15 minutes, with some calls taking as long as an hour. However, in the context of traditional research, where research staff conduct in-person assessments and manually enter data, it remains unclear how much extra time burden this approach actually produces. This learning-by-doing approach may also be difficult to implement depending on research ethics board review timelines. The institutional review board used here enabled rapid review of modifications enabling quick and efficient amendments to the protocol. Finally, this approach may not be scalable for large digital health studies, and it is unclear how effective it might be in other populations.

### Future Directions

Future work should test different aspects of these types of digital engagement approaches in controlled settings, including control groups of financial incentives only and no incentives, to further determine their effectiveness across different populations. Further, the impact (benefits and harms) of returning objective measures of health requires more interrogation. As for increasing understanding of stress detection from wearable-based technologies, other researchers are encouraged to access the Stress and Recovery data that will be hosted on the Synapse at Sage Bionetworks (available in December 2021) to progress this field.

## Conclusions

Digital technologies could facilitate a new era of participant-driven models in research and medicine [56]. As datafication [57], the process of digitizing most aspects of human life, continues to intensify, the need to incorporate insights from the individuals who are the source of those data grows increasingly important. A common shortcoming of the *light touch* digital health study is lacking adequate ground-truth data in the form of participant-reported subjective information. Given the early state of this field, ground truths to validate measurements of health are important yet are so often missing. Although statistical power is important, it is unclear whether an increase in the study size can counter the benefits seen from the depth of collecting additional contextual information. Incorporating a learning-by-doing digital approach facilitates a closed-loop research process whereby participants can offer rich self-reported context for objectively measured data, while researchers can learn in real time and offer knowledge and support back to participants. For this to work, trust and respect between participant and researcher is essential, as it should always be and could serve as a model to be leveraged for the implementation phase of digital medicine.

## Acknowledgments

This work was supported in part, by the Bill & Melinda Gates Foundation (INV-016651). Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to this paper. We would like to thank our protocol officer Dr Andrew Trister for his enabling support. We would like to thank all Stress and Recovery participants for their dedication and contributions to this study, and their efforts on the frontlines during the COVID-19 pandemic.

The following are the Stress and Recovery participants that were included as a group author: Natasha Johnson, Claire Buss, Nadiyah Gary, Toni Gregory, Selena Schmit, Tricia Thomas, Aleisha Lutzen, Brooke Gambino, Klara Oh, Andrea Mizumoto, Jennifer Rand, Rebekah Kreger, Lisa Knudtson, Sarah Landon, Kimberly Ness, Margaret Fitzgibbons, Dee Ann Wolf, MaryAnn G Deiparine, Megan Page, Jeffrey T Richards, Peter McGerty, Lizbeth Rainaud, Bonnie Goff, Brittney Dahlen, Kimberly Fougere, Monica Hoefsmit, Teri Willis, Orange-Marie Miller, Katherine A Linder, Susan Miller, Alexandra L Williams, Natasha Gross, Mindy Churchwell, Whisty Taylor, Lauryn Haigh, Kimberly Whalen, Breanna Stricklin, and Kristine Gerold.

## Authors' Contributions

SMG wrote the manuscript and conducted the analysis. EK assisted with the analysis. SF along with all authors contributed to the design of the study. All authors contributed to the editing of the final manuscript.

## Conflicts of Interest

SF holds <0.05% stock in Oura Health.

Multimedia Appendix 1
Supplemental information.
[PDF File (Adobe PDF File), 525 KB - formative_v5i12e32165_app1.pdf ]

## References

1. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker [accessed 2020-06-05]



XSL•FO
RenderX

2. Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2021-06-06]
3. Global mobile consumer trends: second edition: mobile continues its global reach into all aspects of consumers' lives. Deloitte. 2021 Jun 06. URL: https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-trends.html [accessed 2021-06-06]
4. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the crisis into an opportunity: digital health strategies deployed during the COVID-19 outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]
5. Friend SH. App-enabled trial participation: tectonic shift or tepid rumble? Sci Transl Med 2015 Jul 22;7(297):297ed10 [FREE Full text] [doi: 10.1126/scitranslmed.aab1206] [Medline: 26203077]
6. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21. [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
7. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
8. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
9. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of obtaining measures of lifestyle from a smartphone app: the MyHeart Counts Cardiovascular Health Study. JAMA Cardiol 2017 Jan 01;2(1):67-76. [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]
10. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011. [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
11. Quisel T, Foschini L, Zbikowski SM, Juusola JL. The association between medication adherence for chronic conditions and digital health activity tracking: retrospective analysis. J Med Internet Res 2019 Mar 20;21(3):e11486 [FREE Full text] [doi: 10.2196/11486] [Medline: 30892271]
12. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
13. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
14. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172. [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
15. Andrews JE, Moore JB, Weinberg RB, Sissine M, Gesell S, Halladay J, COMPASS investigators and stakeholders. Ensuring respect for persons in COMPASS: a cluster randomised pragmatic clinical trial. J Med Ethics 2018 Aug;44(8):560-566 [FREE Full text] [doi: 10.1136/medethics-2017-104478] [Medline: 29720489]
16. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756. [doi: 10.1111/acem.12692] [Medline: 25997375]
17. Kaye J, Curren L, Anderson N, Edwards K, Fullerton SM, Kanellopoulou N, et al. From patients to partners: participant-centric initiatives in biomedical research. Nat Rev Genet 2012 Apr 03;13(5):371-376. [doi: 10.1038/nrg3218] [Medline: 22473380]
18. Anderson N, Bragg C, Hartzler A, Edwards K. Participant-centric initiatives: tools to facilitate engagement in research. Appl Transl Genom 2012 Dec 01;1:25-29 [FREE Full text] [doi: 10.1016/j.atg.2012.07.001] [Medline: 24772384]
19. Pruett SB. Stress and the immune system. Pathophysiology 2003 May;9(3):133-153. [doi: 10.1016/s0928-4680(03)00003-8] [Medline: 14567930]
20. Goodday SM, Friend S. Unlocking stress and forecasting its consequences with digital technology. NPJ Digit Med 2019;2:75. [doi: 10.1038/s41746-019-0151-8] [Medline: 31372508]
21. Cabarkapa S, Nadjidai SE, Murgier J, Ng CH. The psychological impact of COVID-19 and other viral epidemics on frontline healthcare workers and ways to address it: a rapid systematic review. Brain Behav Immun Health 2020 Oct;8:100144 [FREE Full text] [doi: 10.1016/j.bbih.2020.100144] [Medline: 32959031]
22. Mozos OM, Sandulescu V, Andrews S, Ellis D, Bellotto N, Dobrescu R, et al. Stress detection using wearable physiological and sociometric sensors. Int J Neural Syst 2017 Mar;27(2):1650041. [doi: 10.1142/S0129065716500416] [Medline: 27440466]
23. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]




XSL•FO
RenderX

24. Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018;1:67. [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]
25. Nosakhare E, Picard R. Toward assessing and recommending combinations of behaviors for improving health and well-being. ACM Trans Comput Healthcare 2020 Mar 02;1(1):1-29. [doi: 10.1145/3368958]
26. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
27. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
28. CIEMS. URL: http://ciems.org/ [accessed 2001-06-06]
29. Active tasks. ResearchKit. URL: http://researchkit.org/docs/docs/ActiveTasks/ActiveTasks.html [accessed 2021-05-01]
30. Penton-Voak IS, Munafò MR, Looi CY. Biased facial-emotion perception in mental health disorders: a possible target for psychological intervention? Curr Dir Psychol Sci 2017 Jun 14;26(3):294-301. [doi: 10.1177/0963721417704405]
31. Cormack F, McCue M, Taptiklis N, Skirrow C, Glazer E, Panagopoulos E, et al. Wearable technology for high-frequency cognitive and mood assessment in major depressive disorder: longitudinal observational study. JMIR Ment Health 2019 Nov 18;6(11):e12814 [FREE Full text] [doi: 10.2196/12814] [Medline: 31738172]
32. Wingenfeld K, Schäfer I, Terfehr K, Grabski H, Driessen M, Grabe H, et al. [The reliable, valid and economic assessment of early traumatization: first psychometric characteristics of the German version of the Adverse Childhood Experiences Questionnaire (ACE)]. Psychother Psychosom Med Psychol 2011 Jan;61(1):e10-e14. [doi: 10.1055/s-0030-1263161] [Medline: 20878600]
33. Murphy A, Steele M, Dube SR, Bate J, Bonuck K, Meissner P, et al. Adverse Childhood Experiences (ACEs) questionnaire and Adult Attachment Interview (AAI): implications for parent child relationships. Child Abuse Negl 2014 Feb;38(2):224-233. [doi: 10.1016/j.chiabu.2013.09.004] [Medline: 24670331]
34. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am J Prev Med 1998 May;14(4):245-258. [doi: 10.1016/s0749-3797(98)00017-8] [Medline: 9635069]
35. Cohen S, Tyrrell DA, Smith AP. Psychological stress and susceptibility to the common cold. N Engl J Med 1991 Aug 29;325(9):606-612. [doi: 10.1056/NEJM199108293250903] [Medline: 1713648]
36. Weathers F, Litz B, Herman D, Huska J, Keane T. The PTSD Checklist (PCL): reliability, validity, and diagnostic utility. 1993 Oct Presented at: Annual Convention of the International Society for Traumatic Stress Studies; October 1993; San Antonio, TX.
37. PTSD CheckList – Civilian Version (PCL-C). MIRECC/CoE. URL: https://www.mirecc.va.gov/docs/visn6/3_ptsd_checklist_and_scoring.pdf [accessed 2021-11-28]
38. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
39. Wagner S, Helmreich I, Dahmen N, Lieb K, Tadic A. Reliability of three alternate forms of the trail making tests a and B. Arch Clin Neuropsychol 2011 Jun;26(4):314-321. [doi: 10.1093/arclin/acr024] [Medline: 21576092]
40. Burghart M, Craig J, Radel J, Huisinga J. Reliability and validity of a motion-based reaction time assessment using a mobile device. Appl Neuropsychol Adult 2019;26(6):558-563 [FREE Full text] [doi: 10.1080/23279095.2018.1469491] [Medline: 30183372]
41. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
42. Karam F, Bérard A, Sheehy O, Huneau MC, Briggs G, Chambers C, OTIS Research Committee. Reliability and validity of the 4-item perceived stress scale among pregnant women: results from the OTIS antidepressants study. Res Nurs Health 2012 Aug;35(4):363-375. [doi: 10.1002/nur.21482] [Medline: 22511354]
43. Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, et al. Development of short forms from the PROMIS™ sleep disturbance and Sleep-Related Impairment item banks. Behav Sleep Med 2011 Dec 28;10(1):6-24 [FREE Full text] [doi: 10.1080/15402002.2012.636266] [Medline: 22250775]
44. Lei DK, Yousaf M, Janmohamed SR, Vakharia PP, Chopra R, Sacotte R, et al. Validation of patient-reported outcomes information system sleep disturbance and sleep-related impairment in adults with atopic dermatitis. Br J Dermatol 2020 Nov;183(5):875-882. [doi: 10.1111/bjd.18920] [Medline: 32017011]
45. Valko PO, Bassetti CL, Bloch KE, Held U, Baumann CR. Validation of the fatigue severity scale in a Swiss cohort. Sleep 2008 Nov;31(11):1601-1607 [FREE Full text] [doi: 10.1093/sleep/31.11.1601] [Medline: 19014080]
46. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
47. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]




XSL•FO
RenderX

48. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
49. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009 Sep;18(7):873-880 [FREE Full text] [doi: 10.1007/s11136-009-9496-9] [Medline: 19543809]
50. McHorney CA, Ware JE, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994 Jan;32(1):40-66. [doi: 10.1097/00005650-199401000-00004] [Medline: 8277801]
51. 36-Item Short Form Survey Instrument (SF-36). RAND Corporation. URL: https://www.rand.org/health-care/surveys_tools/mos/36-item-short-form/survey-instrument.html [accessed 2021-11-28]
52. Oura Ring. URL: https://ouraring.com/ [accessed 2021-01-14]
53. Garmin. URL: https://www.garmin.com/ [accessed 2021-01-15]
54. Russell E, Koren G, Rieder M, Van Uum S. Hair cortisol as a biological marker of chronic stress: current status, future directions and unanswered questions. Psychoneuroendocrinology 2012 May;37(5):589-601. [doi: 10.1016/j.psyneuen.2011.09.009] [Medline: 21974976]
55. IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp; 2020.
56. Erikainen S, Pickersgill M, Cunningham-Burley S, Chan S. Patienthood and participation in the digital era. Digit Health 2019;5:2055207619845546 [FREE Full text] [doi: 10.1177/2055207619845546] [Medline: 31041112]
57. Mejias U, Couldry N. Datafication. Internet Policy Rev 2019;8(4):1-10 [FREE Full text] [doi: 10.14763/2019.4.1428]

## Abbreviations

**REDCap:** Research Electronic Data Capture

*Edited by G Eysenbach; submitted 19.07.21; peer-reviewed by M Eisenstadt; comments to author 07.10.21; revised version received 12.10.21; accepted 27.10.21; published 10.12.21.*

*Please cite as:*
*Goodday SM, Karlin E, Alfarano A, Brooks A, Chapman C, Desille R, Karlin DR, Emami H, Woods NF, Boch A, Foschini L, Wildman M, Cormack F, Taptiklis N, Pratap A, Ghassemi M, Goldenberg A, Nagaraj S, Walsh E, Stress And Recovery Participants, Friend S An Alternative to the Light Touch Digital Health Remote Study: The Stress and Recovery in Frontline COVID-19 Health Care Workers Study*
*JMIR Form Res 2021;5(12):e32165*
*URL: https://formative.jmir.org/2021/12/e32165*
*doi:10.2196/32165*
*PMID:34726607*

©Sarah M Goodday, Emma Karlin, Alexandria Alfarano, Alexa Brooks, Carol Chapman, Rachelle Desille, Daniel R Karlin, Hoora Emami, Nancy Fugate Woods, Adrien Boch, Luca Foschini, Mackenzie Wildman, Francesca Cormack, Nick Taptiklis, Abhishek Pratap, Marzyeh Ghassemi, Anna Goldenberg, Sujay Nagaraj, Elaine Walsh, Stress And Recovery Participants, Stephen Friend. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# The Effect of an Automated Mobile Patient Engagement Application on Emergency Department Revisits: Prospective Observational Study

Pothik Chatterjee[1*], MBA; Adam M Beck[1*], MHS; Jenna Ashley Levenson Brager[1], RN, PhD; Daniel J Durand[1], MD; Christopher R D'Adamo[2], PhD

[1]Innovation and Research Department, LifeBridge Health, Baltimore, MD, United States

[2]Department of Family & Community Medicine, University of Maryland School of Medicine, Baltimore, MD, United States

[*]these authors contributed equally

**Corresponding Author:**
Pothik Chatterjee, MBA
Innovation and Research Department
LifeBridge Health
2401 W. Belvedere Ave
Baltimore, MD, 21215
United States
Phone: 1 410 601 1083
Email: pchatter@lifebridgehealth.org

## Abstract

**Background:** Revisits within 30 days to an emergency department (ED), observation care unit, or inpatient setting following patient discharge continue to be a challenge, especially in urban settings. In addition to the consequences for the patient, these revisits have a negative impact on a health system's finances in a value-based care or global budget environment. LifeBridge Health, a community health system in Maryland, United States, implemented an automated mobile patient engagement application as part of our enterprise-wide digital health strategy to improve patient engagement and reduce revisits to the ED.

**Objective:** The aim of this paper was to evaluate the effectiveness of a customized automated digital patient engagement application (GetWell Loop) to reduce 30-day revisits after home discharge from an ED.

**Methods:** The LifeBridge Health Innovation Department and ED staff from 2 participating health system hospitals collaborated with GetWellNetwork to customize their patient engagement application with automated check-in questions and other on-demand resources (eg, streaming content explaining aspects of self-care during COVID-19). An application link was emailed to adult patients discharged home from the ED. A study of ED visits for patients treated for general medicine and cardiology conditions between August 1, 2018, and July 31, 2019, was conducted using CRISP (Chesapeake Regional Information System for our Patients), Maryland's state-designated health information exchange. We also used data within GetWell Loop (GetWellNetwork) to track patient activation and engagement. The primary outcome was the number of ED patients who experienced a 30-day revisit and who did or did not activate their GetWell Loop account. Secondary outcomes included the overall activation rate and the rate of engagement as measured by the number of logins, alerts, and comments generated by patients through the application. Bivariate analysis comparing outcomes among patients who activated the GetWell Loop application to patients who did not was conducted using the Fisher exact test. Multivariate logistic regression modeling with elastic net regularization was also performed to account for potential confounders and potential collinearity of covariates.

**Results:** During this 1-year study, 1062 (27.4%) of 3866 of all emergency patients treated for general medicine or cardiology conditions, who received an invite to use the digital application, activated their account. The patients discharged from the ED, who were treated for general medicine conditions (n=2087) and who activated their GetWell Loop account, experienced a 30-day revisit rate of 17.3% (n=101) compared with 24.6% (n=369) for those who did not activate their account ($P<.001$). Of the patients treated for cardiology conditions (n=1779), 12.8% (n=61) of those who activated their GetWell account experienced a 30-day revisit compared with 17.7% (n=231) of those who did not activate their account ($P=.01$). The significance of these findings persisted after adjustment for confounding variables including age, race, sex, and payor in logistic regression modeling (adjusted odds ratio 0.75, 95% CI 0.62-0.92; $P=.006$).



XSL•FO
RenderX

**Conclusions:** Our results suggest that a significant percentage of patients are willing to utilize a digital application following ED discharge to better engage in their own care, and that usage of such digital applications may significantly reduce 30-day revisit rates. LifeBridge Health's experience demonstrates that health care systems can leverage automated mobile apps to improve patient engagement and successfully impact clinical outcomes at scale.

***(JMIR Form Res 2021;5(12):e17839)*** doi:10.2196/17839

**KEYWORDS**

patient engagement; value-based care; digital health; mobile app; automation; readmission; revisit; emergency department

## Introduction

Patients who are more actively engaged in their health care experience are more likely to demonstrate better health outcomes and incur lower costs. However, encouraging patients to engage in their health care outside of the clinic can be a challenge. Follow-up communication and adherence to postdischarge instructions are often inconsistent, leading to care gaps and preventable readmissions [1-4].

Revisits within 30 days to an ED, observation care unit, or inpatient setting following patient discharge continues to be a challenge, especially in urban settings. In addition to the consequences for the patient, these revisits have a negative impact on a health care system's finances in a value-based care or global budget environment [5,6].

Mobile apps can help physicians more effectively engage patients outside of acute care settings. Reed et al [2] reported that patients with diabetes and other multiple chronic conditions who connected to health resources via smartphones, tablets, or computers were more likely to be in regular touch with their primary care providers and less likely to be hospitalized. Other groups have demonstrated that offering consumers secure messaging with their providers and the ability to make appointments and view their lab results on mobile devices all led to greater engagement levels [7,8].

Our primary study objective was to evaluate the effectiveness of GetWell Loop to reduce 30-day revisits after home discharge from the emergency department (ED) setting. This study was conducted by the Emergency and Innovation departments of LifeBridge Health, a community health system in Baltimore, Maryland, and implemented at 2 of our hospital EDs—Sinai Hospital of Baltimore and Northwest Hospital Center. Our secondary objectives were to assess patients' willingness to adopt and utilize this technology. Our target adoption rate for the GetWell Loop application was 25% based on the previous efforts by the LifeBridge Health clinical call center as well as the activation and utilization rates of GetWell Loop at other health systems [9-12].

This study examined a digital health intervention to promote continuous patient support and aimed at reducing the number of revisits to the emergency room following discharge. The intervention consisted of a customized version of an already-existing patient engagement application, which was offered to all patients discharged from 2 LifeBridge Health emergency departments. Our hypothesis was that the use of an automated digital patient engagement application would significantly reduce ED revisits as a result of improved patient engagement and education, as well as more frequent check-ins and alerts. GetWell Loop prompts the user to log in to the application to answer questions, receive information, and view content at designated time periods following a specific event. These features are programmed ahead of time and are unique to each individual clinical event. For our study, the clinical event of interest was discharge to home following an ED outpatient encounter.

## Methods

### Design

The intervention studied was a web-based application prompting patients to "check in" for a defined period of time after a visit to the ED. The study population comprised adult patients who returned home from Sinai or Northwest hospitals after an outpatient ED visit related to a well-defined set of conditions described below. The specific web application studied was GetWell Loop. An interdisciplinary team, including both clinical and administrative staff and leaders from Sinai and Northwest hospitals, created an "ED Discharge Care Plan," which included a specific set of questions, resources, and a checklist.

All patients discharged directly home after an ED visit were eligible to use the platform regardless of the reason for treatment. Therefore, the components of the study intervention were designed to apply to all patients treated in the ED with a focus on primary care, given that nearly 25% of the patients treated in the Sinai and Northwest EDs were treated for general medicine and cardiology conditions. To account for this focus, as well as to control for any impact due to a patient's underlying medical condition, the study analysis was limited to those patients treated for general medical and cardiology conditions. In addition, those whose visit led to an inpatient admission were also excluded from this study, as additional modules and content were often added to the intervention in those cases that might further confound the analysis. Figure 1 represents the population enrolled and the specific population analyzed in this study.

The team identified on which postdischarge day each component of the intervention should be delivered to the patient (the cadence), using the discharge date as a reference date (Table 1). The "ED Discharge Care Plan" was designed to stay active for 5 days postdischarge. GetWellNetwork support staff assisted in customizing their software to embed the "ED Discharge Care Plan" in their GetWell Loop platform and automate the delivery of the questions and content through their application. The GetWellNetwork did not fund this study nor were its staff used for any follow up with patients engaging through the platform.



XSL•FO
RenderX

**Figure 1.** Population enrolled on the GetWell Loop application and focus population studied.

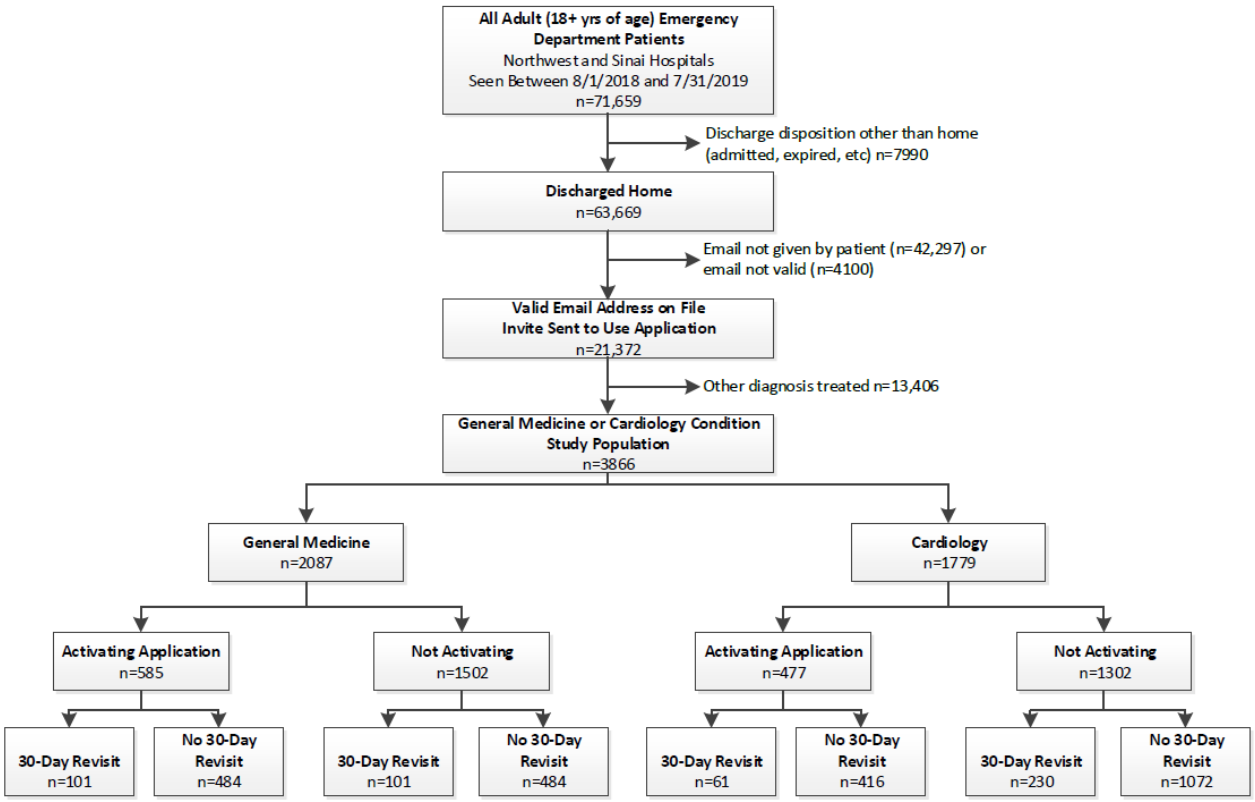



**Table 1.** Check-in questions, resources, and checklists with associated cadence included in the "ED Discharge Care Plan."

| Prompts, resources, and checklists | Days after discharge the questions were scheduled to be sent to the patient |
|---|---|
| **Check-in questions** | |
| Welcome message after ED[a] discharge | 1 |
| If you were given a prescription, do you have any questions about how to take your medications, such as which pills to take or how many times a day? | 1 |
| If you were given a prescription to fill, have you been able to fill it? | 1 |
| Since you visited the emergency department, do you feel your main problem has improved, stayed the same, or worsened? | 1 |
| Do you have any questions about your home care instructions? | 1 |
| Do you have a follow-up appointment? | 1 |
| Do you have any questions about the discharge instructions you received? | 1 |
| Please tell us how satisfied you were with your recent LifeBridge hospital visit? | 2 |
| How satisfied are you with using this application? | 5 |
| **Resources** | |
| Concerning symptoms after ED visit | 1 |
| Managing your follow-up appointments | 1 |
| Taking charge of your medications | 1 |
| **Checklists** | |
| Pick up prescriptions | 1 |

[a]ED: emergency department.

Each question was standardized with a specific answer choice, and each answer choice was identified by the implementation team to trigger a yellow alert, red alert, or no alert. The application also allowed patients to submit free text comments



XSL•FO
RenderX

and questions along with answers to standardized questions. All alerts and comments were highlighted in real time by the GetWell Loop software through a clinical dashboard that was monitored during business days and business hours by agents at LifeBridge Health's clinical call center. Yellow alerts were acted upon within 1 business day, red alerts were handled within an hour, and free-text comments and questions were prioritized individually after triage by the clinical call center staff. The agents in the call center followed up with patients addressing their needs and concerns with a primary focus towards connecting patients with community resources (ie, primary care physician, community pharmacist, etc). Communication with the patient was made either directly through the application or by telephone.

An interface between LifeBridge Health's electronic health record (EHR; Cerner) and the application was implemented in order to auto-enroll patients on the application and initiate the "ED Discharge Care Plan." Enrollment occurred immediately after the patient was discharged. Once enrolled, a user account for the GetWell Loop software was created for each patient, and an invite was emailed the following morning at 5 AM EST, prompting the patient to activate their account and verify their identity. Each patient was required to activate their account after enrollment in order to utilize the application and initiate the automated check-in process. The GetWell Loop software is available to the patient as both a web-based application and as a formal app that may be downloaded on a smartphone.

Once the patient activated their account, the "ED Discharge Care Plan" was initiated, and the first set of check-in questions was presented. The patients continued to receive emails autogenerated by the software to check in based on the configured intervals for each question. Reminder emails were submitted to patients who had not activated their account or missed a check-in. The application is HIPAA (Health Insurance Portability and Accountability Act)-secure and is tied to the patient's EHR number. All communication activity documented by either a care provider or patient within the application was interfaced to the EHR and recorded within the patient's formal electronic medical record.

### Eligibility Criteria

Adult patients (aged 18+ years) discharged home, with a valid email address entered in their record in the EHR, were enrolled. There were no exclusions made based on diagnosis for treatment, patient secondary diagnoses, or any other factor. Patients with an invalid email address or with "bounce backs" to the system were excluded from the study.

### Program Start

The patients were enrolled starting July 18, 2018, and continued through July 31, 2019. All ED staff including physicians, nurses, and registrars were educated with a focus on email collection and informing the patient at discharge. Brochures regarding the software were made available in the ED waiting areas, and a digital poster describing the program was included on all digital displays in the ED.

### Data Extraction and Measurement

The primary intervention was measured by the activation rate defined as the percentage of those invited who activated their account. This information was captured within the GetWell Loop platform.

To analyze the impact on 30-day revisits, CRISP (Chesapeake Regional Information System for our Patients), Maryland's state-designated health information exchange was leveraged. The CRISP database was used to identify those ED visits in which a study patient experienced an emergency department visit, inpatient admission, or observation stay at any Maryland facility within 30 days of discharge. The CRISP database does not include visits to a primary care practice or urgent care facilities. The CRISP reporting system also provided diagnostic category groupings for each ED visit, which allowed the identification of each general medicine and cardiology encounter for analysis. CRISP uses 3M's proprietary Enhanced Ambulatory Patient Grouping System to perform this task.

Information from the EHR was combined with the GetWell and CRISP data. These elements included patients' age, sex, race, primary insurance, and primary diagnosis. The patients' primary insurance was grouped according to the State of Maryland's Health Services Cost Review Commission requirements (ie, Medicare, Medicaid, Medicaid Managed Care, etc).

The date range used for analysis was the 1-year period beginning on August 1, 2018, and ending on July 31, 2019. Any patient who was discharged home and experienced a visit to a Maryland emergency department within 30 days of discharge or who was admitted as an inpatient or observation patient to any Maryland hospital within 30 days of discharge was considered to have experienced a 30-day revisit. The intervention group was defined as those patients who activated their GetWell Loop account and initiated the "ED Discharge Care Plan," regardless of whether they finished all of the modules and regardless of the degree of engagement in the platform.

### Statistical Analysis

Descriptive statistics were computed to characterize the study sample on key demographic characteristics including age, race, sex, and payor status. Bivariate analysis evaluating unadjusted associations between activation of the app and revisit rate was estimated utilizing a Wald chi-square test and a Fisher exact test. Multivariate logistic regression models were also constructed to adjust for potential confounders including age, sex, race, payor status, visit type, and primary diagnosis condition. Elastic net regularization was applied to account for potential collinearity among covariates in the regression models. General medicine and cardiology cohorts were defined based upon the service line groupings in the CRISP data set. Statistical significance was defined as $P<.05$. Statistical analyses were conducted in SAS (version 9.4.1, SAS Institute Inc).

## Results

A total of 3866 patients treated for general medicine (n=2087) and cardiology conditions (n=1779) invited to use GetWell Loop were studied to assess the impact of the application on 30-day revisit rates (Table 2). Combined, this group of patients



XSL•FO
RenderX

represented 3866 (22.4%) of the 17,272 total population invited to use the application and 1062 (24.5%) of the 4337 of the total population who activated their account.

A total of 577 general medicine patients (28% of the total general medicine patients) and 477 cardiology patients (27% of the total cardiology patients) activated their accounts. The average age, sex, and primary payor classification for the cardiology and general medicine patients invited to use the application are listed in Table 2. There was no age difference found between cardiology patients who did not activate their account and those who activated their account (*P*=.64). The age of general medicine patients who activated their account (mean 52 years, SD 18.2) was greater than the age of those who did not activate their account (mean 50 years, SD 18.8; *P*=.02). There was a significantly greater proportion of female patients activating their account in both cardiology and general medicine populations (*P*=.03 and *P*=.02, respectively). A lower proportion of patients with Medicaid insurance activated their account, and there was no difference found in the activation rate for Medicare patients. African American patients activated their account much less frequently (*P*<.001) while White patients activated their account much more frequently (*P*<.001).

**Table 2.** Descriptive statistics for cardiology and general medicine emergency department patients enrolled on GetWell Loop, comparing those activating their account to those not activating their account.

| Patient characteristics | All patients or all conditions invited to use application | All patients or all conditions activating their account | Cardiology patients not activating their account | Cardiology patients activating their account | *P* value[a] | General medicine patients not activating their account | General medicine patients activating their account | *P* value[a] |
|---|---|---|---|---|---|---|---|---|
| Patients, n | 17,272 | 4337 | 1302 | 477 | —[b] | 1502 | 585 | — |
| Age (years), mean (SD) | 45.1 (17.7) | 45.7 (17.4) | 49.5 (16.8) | 49.1 (15.9) | .64 | 49.85 (18.8) | 52.03 (18.2) | .02 |
| Female, n (%) | 11658 (67.5) | 3145 (72.5) | 830 (63.7) | 330 (69.2) | .03 | 955 (63.6) | 404 (69.1) | .02 |
| **Race, n (%)** | | | | | | | | |
| Black or African American | 13430 (77.8) | 3131 (72.2) | 1039 (79.8) | 349 (73.2) | .004 | 1135 (75.6) | 390 (66.7) | <.001 |
| White | 3031 (17.5) | 973 (22.4) | 196 (15.1) | 109 (22.9) | <.001 | 291 (19.4) | 154 (26.3) | <.001 |
| Multiple | 267 (1.5) | 78 (1.8) | 21 (1.6) | 7 (1.5) | .99 | 26 (1.7) | 13 (2.2) | .47 |
| Declined to answer or unknown | 250 (1.4) | 77 (1.8) | 17 (1.3) | 2 (0.4) | .12 | 28 (1.9) | 13 (2.2) | .60 |
| Asian | 178 (1.0) | 51 (1.2) | 19 (1.5) | 9 (1.8) | .52 | 15 (1.0) | 11 (1.9) | .12 |
| American Indian or Alaska Native | 64 (0.4) | 18 (0.4) | 5 (0.4) | 1 (0.2) | .99 | 2 (0.1) | 3 (0.5) | .14 |
| Native Hawaiian, other Pacific Islander | 52 (0.3) | 9 (0.2) | 5 (0.4) | 0 (0) | .99 | 5 (0.3) | 1 (0.2) | .99 |
| **Payor, n (%)** | | | | | | | | |
| Commercial or other payor | 7538 (43.6) | 1927 (44.4) | 546 (41.9) | 249 (52.2) | <.001 | 465 (31.0) | 259 (44.3) | <.001 |
| Medicaid payor | 6141 (35.6) | 1301 (30.0) | 377 (29.0) | 117 (24.5) | .07 | 500 (33.3) | 130 (22.2) | <.001 |
| Medicare payor | 3593 (20.8) | 849 (19.6) | 311 (23.1) | 96 (20.1) | .10 | 444 (29.6) | 170 (29.1) | .82 |

[a]*P* value calculated using the Welch 2-sample *t* test (age), chi-square test (sex), and Fisher exact test (payor, race).

[b]Not applicable.

A multivariate logistic regression model was constructed to assess the impact of the intervention while adjusting for age, sex, race, primary payor, and primary diagnosis condition (general medicine or cardiology). After adjustment for covariates, patients who activated the application were significantly less likely to have a 30-day revisit (odds ratio 0.75, 95% CI 0.62-0.92; *P*=.006).

Age, sex, and race were not found to have a significant impact on the 30-day revisit outcome. Medicaid and Medicare patients were more likely to have a 30-day revisit (*P*<.001 and *P*=.02, respectively). The patients treated for a cardiology condition compared with those treated for general medicine were found to be less likely to have a 30-day revisit rate (*P*<.001).

The 30-day revisit rate for the general medicine subgroup who activated their GetWell Loop account was 17.3% (n=101) compared to 24.6% (n=369) of those who did not activate their account (*P*<.001; Table 3). For cardiology patients who activated their account, 12.8% experienced a 30-day revisit compared to 17.7% (n=231) of those who did not activate their account (*P*=.01). The percentage of check-in questions triggering the alert were as follows: follow-up appointment assistance (48%, n=917), prescription fill assistance (16%, n=306), discharge



XSL•FO
RenderX

instruction questions (12%, n=229), understanding of treatment plan (10%, n=191), health status worsening (9%, n=172), med instruction questions (5%, n=96).

**Table 3.** Number of patients (and the associated 30-day revisit rates) treated for a general medical or cardiology condition in the Emergency Department, who were invited to use the GetWell Loop application.

| Application usage or nonusage | Revisit within 30 days of discharge home, n | No revisit within 30 days of discharge home, n | Total, n | 30-day revisit rate (%) | *P* value |
|---|---|---|---|---|---|
| **Use of the application: general medicine patients** | | | | | <.001 |
| Number of patients not activating their account | 370 | 1132 | 1502 | 24.6 | |
| Number of patients activating their account | 101 | 484 | 585 | 17.3 | |
| Total | 471 | 1616 | 2087 | 22.6 | |
| **Use of the application: cardiology patients** | | | | | .01 |
| Number of patients not activating their account | 230 | 1072 | 1302 | 17.7 | |
| Number of patients activating their account | 61 | 416 | 477 | 12.8 | |
| Total | 291 | 1488 | 1779 | 22.6 | |

# Discussion

## Strengths

The study benefitted from the use of many specific EHR data fields that are required for registration or for state reporting. Age, sex, race, primary payor, discharge disposition, and principal diagnosis are required state reporting elements and are reviewed at registration and during coding and billing. The email field is a required field during registration and includes an option for “none” or “patient refused.” These fields have been required elements for many years, and there were no new fields introduced in the EHR for this study.

The study also benefitted from automated enrollment on the platform based on these data fields. We did not have to rely on human intervention to specifically identify subjects for the study, meaning that any data-entry error and bias was applied across all potential candidates. All patient participants were invited to use the application in the same manner regardless of age, race, or socioeconomic status. We also limited our analysis to those patients who provided a valid email and thus invited to use the application rather than including all adult patients who returned home from their ED visit.

Our organization’s participation in Maryland’s health information exchange (CRISP) along with all other health systems in the state provided a robust and uniquely comprehensive data set for 30-day revisits. For several years, CRISP has been used to develop a model that links patient activity across all hospitals and their associated inpatient, ED, and observation care units leveraged for the analysis.

## Limitations

There were several limitations to our study. Patient acuity, secondary diagnoses, and other comorbid and socioeconomic circumstances play a role in a given patient’s potential to revisit an ED or to be readmitted to a hospital within 30 days after discharge from an ED. The study controlled for acuity by analyzing only adult patients treated for similar conditions (cardiology and general medicine) with a similar discharge disposition (discharged home directly from the ED). Cardiology and general medicine conditions were based on clinical groupings, and there was no attempt to analyze per the 10th revision of the International Classification of Diseases or to factor in secondary conditions. In an effort to control for technology use by patients and their likelihood of using the application, the study focused only on those patients who were invited to use the application and thus had provided a valid email address at registration.

There was no analysis of or comparison to those patients who were not invited, nor was there any means to control for the patients’ access to the internet, a mobile device, or a computer that is required for using the application. During the study, there was only 1 known organization-wide effort that was introduced that might have impacted the results. The community care coordinators in the EDs at both hospitals in this intervention were engaging identified patients who were high utilizers of ED services and were following up with these patients by telephone after discharge to the community; this was a small percentage of the total ED population, and it is felt to be unlikely that this significantly impacted the results discussed above. Our study was not able to determine whether this subgroup was enrolled on the application, and if so, whether they activated their account. Our study was also not able to identify and control for any other characteristic differences between those patients who activated their account and those who did not, such as education, lifestyle, social determinants of health (housing, transportation, and substance abuse), and access to primary care providers.

## Conclusion

The study revealed a statistically significant association between the use of the digital application and a lower revisit rate after discharge home. These results indicate the potential value of digital health applications to improve 30-day revisit rates.

The relative 30% lower revisit rate across both general medical and cardiology conditions sends a strong signal that the adoption of digital patent engagement tools can improve specific population-health outcomes and warrants further analysis to control for potential selection bias and chronic or comorbid conditions that may have additional patient acuity, which in



XSL•FO
RenderX

turn may have impacted this study. The results also demonstrate that a significant percentage of patients are willing to utilize web-based applications to proactively engage in their own care following discharge. LifeBridge Health's experience demonstrates that health care systems can leverage automated mobile apps to improve patient engagement and successfully impact clinical outcomes at scale. Further research should focus on expanding clinical use cases, enhancing activation rates, and studying and addressing barriers to patient adoption (eg, the impact of social determinants of health) in order to ensure that these methods improve, rather than exacerbate, disparities in health outcomes.

## Conflicts of Interest

Effective January 1, 2021, DJD began a term as a formal advisor to GetWellNetwork. All revenues from this relationship go to LifeBridge Health, and this relationship was established after the conclusion of the study in question.

## References


1. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020 Nov 01;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
2. Reed ME, Huang J, Brand RJ, Neugebauer R, Graetz I, Hsu J, et al. Patients with complex chronic conditions: Health care use and clinical events associated with access to a patient portal. PLoS One 2019;14(6):e0217636 [FREE Full text] [doi: 10.1371/journal.pone.0217636] [Medline: 31216295]
3. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: A systematic, narrative review. Patient Educ Couns 2019 Jun;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
4. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push Notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
5. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 2009 Apr 02;360(14):1418-1428. [doi: 10.1056/NEJMsa0803563] [Medline: 19339721]
6. Report to the Congress: Promoting Greater Efficiency in Medicare. Washington, DC: Medicare Payment Advisory Commission; 2007. URL: http://www.medpac.gov/docs/default-source/reports/Jun07_EntireReport.pdf [accessed 2021-11-04]
7. Bach RL, Wenz A. Studying health-related internet and mobile device use using web logs and smartphone records. PLoS One 2020;15(6):e0234663 [FREE Full text] [doi: 10.1371/journal.pone.0234663] [Medline: 32530937]
8. James J. Patient Engagement. Health Affairs - Health Policy Brief. 2021 Sep 21. URL: https://www.healthaffairs.org/do/10.1377/hpb20130214.898775/full/ [accessed 2021-11-23]
9. Pollack A, Backonja U, Miller A, Mishra SR, Khelifi M, Kendall L, et al. Closing the Gap: Supporting Patients' Transition to Self-Management after Hospitalization. 2016 May Presented at: Proc SIGCHI Conf Hum Factor Comput Syst; 2016; San Jose, California, USA p. 5324-5336 URL: http://europepmc.org/abstract/MED/27500285 [doi: 10.1145/2858036.2858240]
10. Arora S, Burner E, Terp S, Nok Lam C, Nercisian A, Bhatt V, et al. Improving attendance at post-emergency department follow-up via automated text message appointment reminders: a randomized controlled trial. Acad Emerg Med 2015 Jan;22(1):31-37 [FREE Full text] [doi: 10.1111/acem.12503] [Medline: 25388481]
11. Shuen JA, Wilson MP, Kreshak A, Mullinax S, Brennan J, Castillo EM, et al. Telephoned, texted, or typed out: a randomized trial of physician-patient communication after emergency department discharge. J Emerg Med 2018 Oct;55(4):573-581 [FREE Full text] [doi: 10.1016/j.jemermed.2018.07.023] [Medline: 30181075]
12. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]


## Abbreviations

**CRISP:** Chesapeake Regional Information System for our Patients
**ED:** emergency department
**EHR:** electronic health record



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 15.01.20; peer-reviewed by J Cohen, J Zulueta; comments to author 03.09.20; revised version received 20.01.21; accepted 06.07.21; published 13.12.21.*

*Please cite as:*
*Chatterjee P, Beck AM, Brager JAL, Durand DJ, D'Adamo CR*
*The Effect of an Automated Mobile Patient Engagement Application on Emergency Department Revisits: Prospective Observational Study*
*JMIR Form Res 2021;5(12):e17839*
*URL: https://formative.jmir.org/2021/12/e17839*
*doi:10.2196/17839*
*PMID:34898451*

©Pothik Chatterjee, Adam M Beck, Jenna Ashley Levenson Brager, Daniel J Durand, Christopher R D'Adamo. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Integration of Web Analytics Into Graduate Medical Education: Usability Study

Jackson Massanelli[1], BSc; Kevin W Sexton[2], MD; Chris T Lesher[1], BSc; Hanna K Jensen[2], MD, PhD; Mary K Kimbrough[2], MD; Anna Privratsky[2], DO; John R Taylor[2], MD; Avi Bhavaraju[2], MD

[1]College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR, United States

[2]Division of Trauma and Acute Care Surgery, Department of Surgery, University of Arkansas for Medical Sciences, Little Rock, AR, United States

**Corresponding Author:**
Avi Bhavaraju, MD
Division of Trauma and Acute Care Surgery
Department of Surgery
University of Arkansas for Medical Sciences
4301 W Markham St
Little Rock, AR, 72205
United States
Phone: 1 501 320 7233
Email: avi99@uams.edu

## Abstract

**Background:** Web analytics is the measurement, collection, analysis, and reporting of website and web application usage data. While common in the e-commerce arena, web analytics is underutilized in graduate medical education (GME).

**Objective:** The University of Arkansas for Medical Sciences Department of Surgery website was revamped with input from in-house surgeons in August 2017. This study investigated the use of web analytics to gauge the impact of our department's website redesign project.

**Methods:** Google Analytics software was used to measure website performance before and after implementation of the new website. Eight-month matched periods were compared. Factors tracked included total users, new users, total sessions, sessions per user, pages per session, average session duration, total page views, and bounce rate (the percentage of visitors who visit a site and then leave [ie, bounce] without continuing to another page on the same site).

**Results:** Analysis using a nonpaired Student *t* test demonstrated a statistically significant increase for total page views (before vs after: 33,065 vs 81,852; $P<.001$) and decrease for bounce rate (before vs after: 50.70% vs 0.23%; $P<.001$). Total users, new users, total sessions, sessions per user, and pages per session showed improvement. The average session duration was unchanged. Subgroup analysis showed that after the main page, the next 3 most frequently visited pages were related to GME programs in our department.

**Conclusions:** Web analytics is a practical measure of a website's efficacy. Our data suggest that a modern website significantly improves user engagement. An up-to-date website is essential for contemporary GME recruitment, will likely enhance engagement of residency applicants with GME programs, and warrants further investigation.

***(JMIR Form Res 2021;5(12):e29748)*** doi:10.2196/29748

**KEYWORDS**

graduate medical education; website analysis; residency recruitment; medical education; website; analytics; usage; usability; user engagement; user-centered design; website design

## Introduction

Web analytics is the measurement, collection, analysis, and reporting of website and web application usage data. Web analytics has been used in many industries [1-3] to understand and optimize their websites by improving user engagement and stimulating traffic [2-5]. These tools facilitate the measurement of a website's efficacy using metrics, such as the bounce rate, which is the percentage of users who visit a site and then leave (ie, bounce) without continuing to another page on the same site [6]. The bounce rate represents a website's effectiveness in encouraging users to continue their visit beyond the first page



XSL•FO
RenderX

[7]. A high bounce rate implies that users are uninterested in the site's content or that the design is unsuitable for the user [8]. The bounce rate and other web analytical tools are underutilized in graduate medical education (GME).

The bounce rate and other similar web metrics are commonly utilized in e-commerce and web-marketing to identify user trends, improve website flow and usability, and ultimately boost web traffic. Examples of the contemporary use of web analytics would be an e-commerce website measuring user traffic during a sale or a holiday to see which items garner the most attention, a software company using page metrics to direct visitors to help forums or FAQ sections faster, an educational website tracking user data to more easily provide information on admissions or courses to its students, and a political campaign measuring a candidate's web traffic after a debate [9].

A high bounce rate is concerning for a site because it means visitors do not browse through the entire site and do not get a comprehensive look at all it has to offer [6], which generally indicates that the site or page is not relevant to its visitors. The bounce rate has been described as the measure of a user's satisfaction with a given page [10], and most businesses agree that pages with high bounce rates should be reworked to be more user-friendly, with the goal of making a user's visits more frequent and longer in duration, thus lowering the bounce rate [3]. Similarly, lowering the bounce rate is generally associated with improved user engagement by allowing the user to explore more content in a pleasing and persuasive fashion [3]. Reciprocally, studies have also demonstrated that higher user satisfaction lowered the bounce rate [11]. A high bounce rate could be due to a multitude of factors, but is primarily attributed to an illogical navigation scheme, irrelevant and disorganized content, or faulty design [3]. Our website redesign project aimed to address all of these issues.

It is well established that residency applicants find a GME program's website a valuable tool in determining where to apply, where to interview, and how to rank individual programs [12]. In 2003, Mahler et al showed that 96.5% of emergency medicine applicants used the web as a resource to investigate residency programs, and almost all of the applicants applied to programs they thought had the best website [13]. Many individual residency and specialty programs have subsequently evaluated their own websites, creating a vast amount of literature on the subject [12-26]. These studies add to the overall subject of human-computer interaction (HCI), where web analysis has been studied extensively [27]. These studies found that while many applicants use a program's website to gather information, the sites themselves lack all of the content items applicants are looking for (eg, faculty research and resident biographies), leaving room for improvement [12-26]. The purpose of our website is multifold. It gives users an introduction to our department, lists the organizational infrastructure, provides information about faculty and their research endeavors, acts as a recruitment tool by providing demographic info about the program and its residents, serves as a resource for undergraduate students, and, in some cases, provides clinical practice guidelines.

Our study investigated the use of web analytics to gauge the impact of our institution's Department of Surgery website redesign project, which was done in close collaboration with in-house surgeons. While website redesign has been extensively studied in HCI research for the last 30 years [27], it does not seem to be as extensively studied or used in the GME setting. To our knowledge, no previous study has used analytical data to evaluate a GME website and very few [18,24] completed a follow-up assessment after redesigning their website to determine if changes to the website were effective. The purpose of this study was to utilize web analytics to gauge the impact of our department's website redesign project and prove with objective metrics that a fresh and up-to-date website will improve web traffic for a GME website and thus will improve a GME program's exposure to prospective applicants and will potentially help with residency recruitment. The objectives of this study were to (1) examine and analyze user data for our department website before and after the redesign and (2) assess if the implemented changes improved user engagement, web traffic, and throughput. The goal of our project was to determine if a fresh and modern GME website can be used as an effective recruiting tool.

## Methods

The Department of Surgery website for the University of Arkansas for Medical Sciences (UAMS) was redesigned during the summer of 2017 and launched on August 30, 2017. Using Google Analytics software, multiple variables were tracked to gauge the website's performance. Google Analytics is a set of free online tools that track website visitors' mouse clicks and information requests. The data gathered by Google Analytics can be used to see which pages on an organization's website are the most popular or most accessed, what type of information visitors are interested in accessing, what path visitors take as they navigate to and away from an organization's website, how much time they spend on the site, and a variety of other metrics [9]. These data are not stored on visitors' computers and contain no personally identifiable information. Google Analytics works by having a website administrator attach code onto each server that hosts the web page, and begins tracking the page as soon as it is uploaded [9]. The Google Analytics code is easy to upload to a website's server, and Google offers step-by-step instructions to facilitate implementation, making it an easily accessible option for any program that wants to track user or site metrics [9]. For our study, we examined monthly and aggregate user and webpage data to examine the effects of the redesign project. Data were compared for 8-month matched periods from January to August of 2017 and 2018, representing data from before and after the new site was launched, respectively. The months of September through December for each year were excluded to account for the seasonal effect of GME interviews adversely elevating website traffic.

The study population included all visitors to the website during the aforementioned time periods. No users were excluded. Users remained anonymous and were not individually tracked, and all collected data were completely deidentified.



XSL•FO
RenderX

The primary end-points examined were total page views and bounce rate. Secondary variables included total users, new users, total sessions, sessions per user, pages per session, and average session duration. The Google Analytics software package also provided a breakdown of the activity on each page of the website, allowing us to stratify which pages were the most heavily trafficked, and thus, which sections of the new website were the most popular among our users. An unpaired Student *t* test was used to test for significant differences among our variables. Continuous variables were evaluated using GraphPad Prism Version 8.0.0 (GraphPad Software, Inc). The project received a nonhuman subject research exemption from the UAMS Institutional Review Board (IRB #249934).

## Results

After the departmental website redesign, improvements were observed in most of the tracked metrics, including both our primary end-points. A statistically significant increase in total page views ($P<.001$; Figure 1) and a statistically significant decrease in bounce rate ($P<.001$; Figure 2) were observed. There were also statistically significant improvements in the number of total users ($P<.001$; Figure 1), number of new users ($P<.001$; Figure 1), number of sessions ($P<.001$; Figure 1), average number of sessions per user ($P=.004$; Figure 3), and pages viewed per session ($P<.001$; Figure 3). The average session duration was unchanged. The chronological breakdown of total monthly page views for the 8-month matched periods preupdate and postupdate are shown in Figure 4.

**Figure 1.** Graph demonstrating significant increases in total users, new users, total sessions, and page views after the website update.

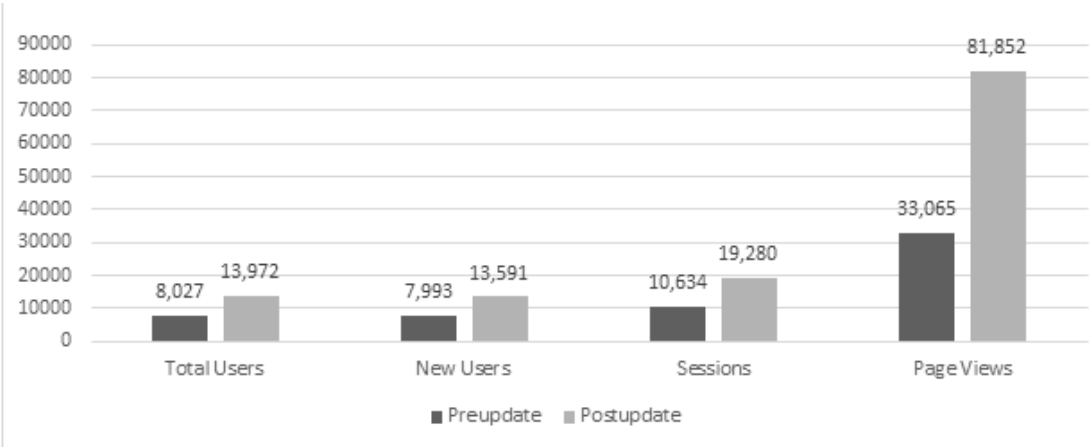
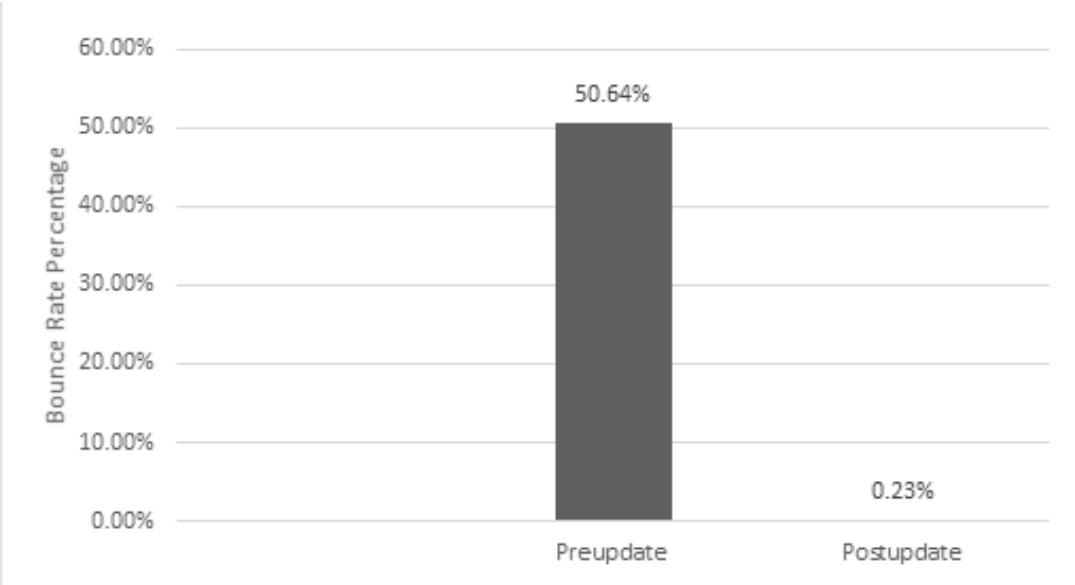



**Figure 2.** Graph demonstrating a significant decrease in the bounce rate after the website update.






XSL•FO
RenderX

**Figure 3.** Graph demonstrating an increase in pages per session and similarity between sessions per user.

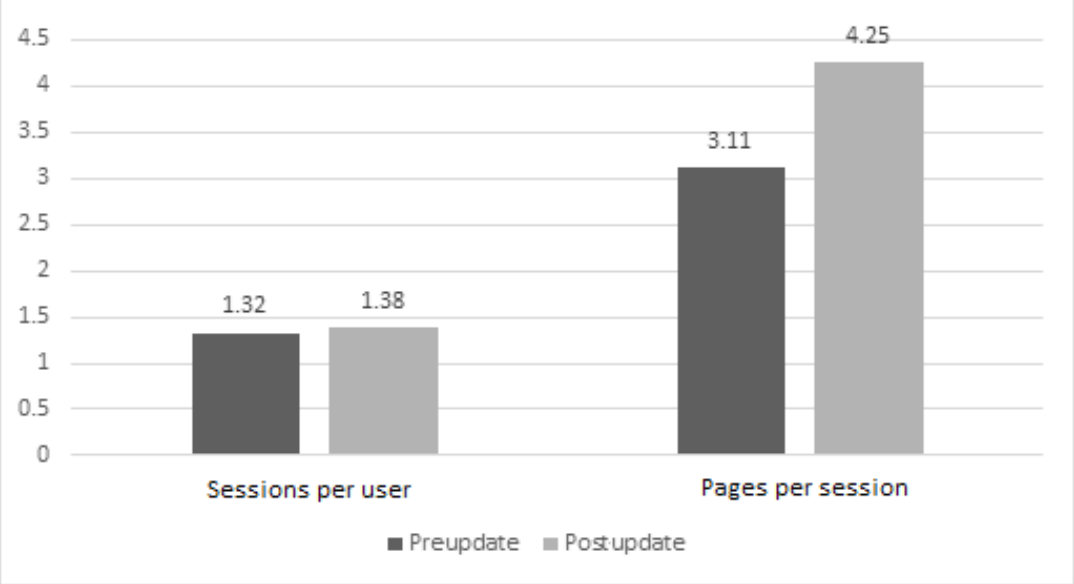
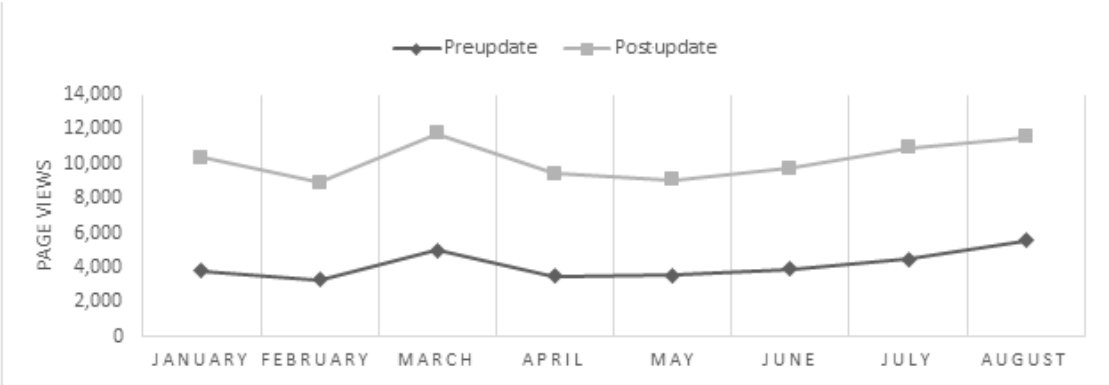



**Figure 4.** Graph comparing page views between 8-month matched periods before and after the website update.




Before the update, the top 5 most visited pages were (1) Home Page, (2) Our Residents, (3) General Surgery Residencies and Fellowships, (4) Residencies and Fellowships, and (5) Vascular Surgery. After the redesign, the top 5 most visited pages were (1) Our Residents, (2) Junior Clerkship Goals and Objectives, (3) Home Page, (4) Trauma Guidelines, and (5) Residencies and Fellowships.

## Discussion

### Principal Findings

After modernizing our department website, we noted significant improvements in virtually all tracked metrics. Utilizing Google Analytics, we were able to show a nearly 2.5-fold increase in the total number of page views and a statistically significant drop in our website's bounce rate from 50.64% (N=10,634) to 0.23% (N=19,280) when comparing 8-month matched periods before and after the updated website was launched. Secondary analysis of the user data showed that the most commonly visited pages on the website were related to our undergraduate and graduate medical education programs.

Our previous department website received universally poor feedback for being outdated, incomplete, and difficult to navigate. Based on an institution-wide initiative to update the UAMS web presence, we performed a total overhaul of the Surgery Department website to improve the organization, flow, usability, and multimedia component of the site through the addition of pictures and videos. The principles used to guide the redesign were related to optimization of web content, frequent and consistent content updates, improving content accessibility and responsiveness, search engine optimization, and maintaining institutional branding. Overall, the new website's content follows the Web Content Accessibility Guidelines (WCAG) 2.1 standard, which provides guidelines for the way content is presented on a website (eg, the way graphics and colors are used, column width, moving elements, text size, etc). Our website redesign plan allowed us to switch website templates to the new standard, which also presented an opportunity to update its content. Responsiveness is the ability of a site to adjust to different resolutions and devices in real time, for example, zooming in and out of the website on a mobile device. Older websites, including our previous website, are static and thus unable to accommodate for these differences, leading to a less favorable user experience. Our updated website, using the new WCAG 2.1 standards, is much better able to handle these differences, providing a more fluid user experience.



XSL•FO
RenderX

Search engine optimization is the process of maximizing the number of visitors to a particular website by ensuring that the site appears high on the list of results returned by a search engine, such as Google and Yahoo. Search engine optimization is affected by the code of the template used to create the website, specific webpage content, and site organization. Branding refers to the overall theme used on a given site. Newer web templates are more stringent about the use of colors and logos in order to keep things consistent, reduce variation from page to page, and maintain the brand throughout a given website.

Some of the specific changes we made included the following: a detailed breakdown of each division within the department, including a comprehensive list of faculty within each division with titles and links to clinical bios; descriptions of each division's current priorities, initiatives, and research endeavors with specific divisional guidelines in some cases; and a dedicated section for medical students, with detailed information on the M3 junior clerkship, M4 surgery electives, M4 surgery honors course, and Summer in Surgery program, an innovative program geared toward M1 students interested in pursuing a career in surgery. Regular prospective updates are made to the new site every few months to ensure the information is up-to-date, including the addition of new faculty members, faculty research endeavors, new clinical and operational guidelines, department news, and divisional updates.

While there is a multitude of research in HCI analyzing web traffic, there is a paucity of research in examining GME websites, especially in medical education literature. Much of the research evaluating GME websites up to this point has been qualitative, comparing a particular site to a predetermined checklist of features, either chosen by the investigators [14-25] or suggested by residency applicants [12,13,26]. Little work has been done to investigate how improvements to a program's website impact the user experience or applicant's perceptions of a program, and to the best of our knowledge, no studies have taken a quantitative data-driven approach to this process. Web analytic software provides GME programs the means to objectively evaluate their websites and the effect any improvements may have on end-users, such as residency applicants. Prior studies have demonstrated that Google Analytics is an effective tool to assess web traffic on health sites. Pang et al used Google Analytics to assess desktop and mobile website traffic for a health channel website. They examined traffic over a 3-month period and found that the most heavily visited pages were not the home page and search results like they hypothesized, concluding that website owners should examine their own web traffic to tailor their designs to their users' diverse needs, search approaches, and behaviors [28]. They also examined users' outgoing pages, which varied from the home/search page to other specific content pages, but they were not able to determine why they left the site on those pages (eg, the user found all the information, the user is going to look on another site, etc) [28].

Prior studies have evaluated GME websites in almost every specialty [12-25]. Embi et al published one of the first studies to show a program's website is a critical tool for GME recruitment [12]. They surveyed a large cohort of residency applicants to a single internal medicine program and determined that internet-based tools were more viable options for information delivery than paper resources, such as printed brochures. They reported that 80% of applicants used websites to help decide where to apply, 69% used websites to determine where to interview, and 36% used websites to help rank order programs for The National Resident Matching Program [12]. Mahler et al surveyed emergency medicine residency applicants and showed that online information from programs' websites influences an applicant's decision by highlighting aspects of the program that are pertinent to the applicant and making them easy to find; 40% of applicants rated an easily navigated website as "very" or "moderately important" to their decision-making process [13]. They hypothesized that an easily navigated complete website may improve recruitment to emergency medicine residency programs.

Our study found that after the redesign, there was a significant increase in web traffic to the website. Google Analytics software showed improvements in virtually all tracked metrics, including total page views, total users, new users, sessions, sessions per user, pages viewed per session, and most importantly bounce rate. We suspect that the improved metrics of the new site are due to better flow and organization, ease of navigation, and being much more user friendly. The significance of this tremendous drop in the bounce rate cannot be understated. Prior to the redesign, approximately 50% of users visited the first page and immediately left ("bounced"), not interacting with any other portion of the site, and after the update, only 0.23% of all users bounced, meaning 99.77% of users visited a subsequent page on the site. While we cannot directly attribute the improved website metrics to any particular change and, at present, cannot directly tie the improved web traffic to improved residency recruitment, based on feedback about the old site from students, residents, and faculty, we postulate that the new website's ease of use and fresh look played a very direct role in improving user experience and user engagement, as evidenced by the tremendous drop in the bounce rate. We hypothesize that this improved user engagement and web traffic will ultimately translate downstream to improved GME recruitment, which is the focus of the next phase of this project.

Based on the above data, we believe the implications for GME are clear. As e-commerce sites must attract new users (ie, shoppers) to make sales, GME programs must make themselves attractive to residency applicants (ie, "the consumer"). The program or department website is often the first contact a visitor has with a program and its first opportunity to sell itself to the prospective applicant; thus, having a well-designed and visually appealing website is essential to making an impactful first impression, particularly for someone who is not familiar with the program. An easy-to-use website with an abundance of information applicants are looking for gives a program credibility and makes it more appealing to the "buyer" (ie, the applicant). We predict this positive first impression could help persuade applicants to think more favorably about a program, particularly those they may not have otherwise considered. We hypothesize that programs with a modern website will improve their web metrics, which will translate to increased engagement of users with these programs and enhanced opinions of the programs by applicants. For programs that focus attention on



XSL•FO
RenderX

their web presence and utilize web analytics in this way, we theorize this will ultimately lead to improved match rates and matching of preferred applicants. Based on our experience, we believe that using web analytics to track user data is an important but underutilized tool that can help GME programs identify areas of weakness and optimize their web presence to make themselves more marketable to prospective GME applicants. Without a focused and deliberate effort in this regard, contemporary GME recruitment is inadequate. We believe that a program with a modern website that focuses attention on web presence and utilizes web analytics will improve metrics, which will translate to increased user engagement, enhanced opinions of these programs by applicants, and ultimately improved match rates and matching of preferred applicants. With respect to our own GME program, the numbers of applicants to our program before the update were 773 (2016) and 996 (2017). After the update, we had 793 (2018), 871 (2019), and 1059 (2020) applicants (unpublished data). While there was no immediate boost in the number of applicants after the update, the number of applicants has steadily increased since then. This increase could be due to a number of factors, many of which are outside the scope of this study, but based on the improved web metrics of the new website and the results from a follow-up study we conducted surveying applicants to our program about their opinions and experiences with the new website, we firmly believe the new website played a positive role in driving applicant numbers up. In that study (N=121), the results of which have not yet been published, the majority (98.3%) of interviewees visited our department website before their interview day, the highest-rated features of our site were easy navigability and clean design (57.9%), the majority (64.5%) of applicants reported that the department's website influenced their opinions of a program, and the majority (94.2%) of applicants reported that a well-developed user-friendly department website is an important factor in selecting a residency program (unpublished data).

In a previous study that assessed the quality of general surgery residency program websites for accessibility, ease of use, design, and content, among the 167 program websites evaluated, an average of 6 out of 16 content items were present and 6 out of 10 design principles were followed, when compared to established principles of website design and content [21]. To address this concern with our site, as part of the secondary analysis, we examined our usage data to track which subpages had the heaviest traffic and lowest bounce rates. Prior to the redesign, the most popular page on the site was the "Home Page," followed by "Our Residents," "General Surgery Residency and Fellowships," "Residencies and Fellowships," and "Vascular Surgery." After the update, the most popular pages were "Our Residents," "Junior Clerkship Goals and Objectives," "Home Page," "Trauma Guidelines," and "Residencies and Fellowships." Based on this analysis, it is readily apparent that the majority of visitors to the site are looking for information about our undergraduate and graduate medical education programs, which allowed us to tailor the site to the needs and desires of our users. The improved usability and updated information have helped increase traffic 2.5 fold, which we hypothesize will improve our ability to attract and recruit residency applicants. Previous studies have shown that what applicants prioritize on a program's website varies between specialties, so we suggest that GME programs should survey their applicant pool to determine where to focus their attention and use web analytic tools to follow user traffic and bounce rates. By doing so, they can optimize their web presence and leverage their department and program websites as powerful recruiting tools to attract prospective applicants. For all these reasons, we feel strongly that an organized and up-to-date department or program website is an essential component of contemporary GME recruitment.

## Limitations

One limitation of our study is that we did not directly survey visitors to gauge their opinions about our website in real time. While there were significant improvements in virtually all of our tracked data points after the update, which we attribute to the improved website design, this study did not determine which specific changes were responsible for the improved traffic. We also cannot say with certainty that the changes to the website were responsible for the improved user traffic, but we have no other plausible explanation for this other than the updated website.

## Conclusion

Web analytics is a quick and practical measure of a website's efficacy. Our data suggest that a fresh and modern website significantly improves user engagement, web traffic, and throughput. Based on these data, we hypothesize that a polished and up-to-date website is essential for contemporary GME recruitment, will enhance engagement of residency applicants with GME programs, and warrants further investigation.

## Conflicts of Interest

None declared.

## References

1. Berman B. Planning and implementing effective mobile marketing programs. Business Horizons 2016 Jul;59(4):431-439. [doi: 10.1016/j.bushor.2016.03.006]
2. Chaffey D, Patron M. From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. J Direct Data Digit Mark Pract 2012 Aug 8;14(1):30-45. [doi: 10.1057/dddmp.2012.20]
3. Vendivel M. Virtual Rebel Website: A Strategy to Increase User Engagement through Bounce Rate Analysis. UNLV Theses, Dissertations, Professional Papers, and Capstones. 2014. URL: https://digitalscholarship.unlv.edu/thesesdissertations/2154/ [accessed 2021-11-19]



XSL•FO
RenderX

4. Hasan L, Morris A, Probets S. Using Google Analytics to Evaluate the Usability of E-commerce Sites. In: Kurosu M, editor. Human Centered Design. HCD 2009. Lecture Notes in Computer Science, vol 5619. Berlin, Heidelberg: Springer; 2009:697-706.
5. Gofman A, Moskowitz HR, Mets T. Integrating science into web design: consumer‐driven web site optimization. Journal of Consumer Marketing 2009 Jun 26;26(4):286-298. [doi: 10.1108/07363760910965882]
6. Arendt J, Wagner C. Beyond description: converting web site usage statistics into concrete site improvement ideas. Journal of Web Librarianship 2010 Mar 04;4(1):37-54. [doi: 10.1080/19322900903547414]
7. Common Language in Marketing. Marketing Accountability Standards Board. URL: https://themasb.org/projects/current/common-language/ [accessed 2020-09-20]
8. Zheng G. Web Analytics Overview. In: Encyclopedia of Information Science and Technology, Third Edition. Hershey, PA: IGI Global; 2015.
9. Kent ML, Carr BJ, Husted RA, Pop RA. Learning web analytics: A tool for strategic communication. Public Relations Review 2011 Dec;37(5):536-543. [doi: 10.1016/j.pubrev.2011.09.011]
10. Egri G, Bayrak C. The role of search engine optimization on keeping the user on the site. Procedia Computer Science 2014;36:335-342. [doi: 10.1016/j.procs.2014.09.102]
11. Sculley D, Malkin R, Basu S, Bayardo R. Predicting Bounce Rates in Sponsored Search Advertisements. In: KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009 Presented at: 15th ACM SIGKDD international conference on Knowledge discovery and data mining; June 28, 2009-July 1, 2009; Paris, France p. 1325-1334. [doi: 10.1145/1557019.1557161]
12. Embi PJ, Desai S, Cooney TG. Use and utility of Web-based residency program information: a survey of residency applicants. J Med Internet Res 2003 Sep 25;5(3):e22 [FREE Full text] [doi: 10.2196/jmir.5.3.e22] [Medline: 14517113]
13. Mahler SA, Wagner M, Church A, Sokolosky M, Cline DM. Importance of residency program web sites to emergency medicine applicants. J Emerg Med 2009 Jan;36(1):83-88. [doi: 10.1016/j.jemermed.2007.10.055] [Medline: 18439790]
14. Huang BY, Hicks TD, Haidar GM, Pounds LL, Davies MG. An evaluation of the availability, accessibility, and quality of online content of vascular surgery training program websites for residency and fellowship applicants. J Vasc Surg 2017 Dec;66(6):1892-1901 [FREE Full text] [doi: 10.1016/j.jvs.2017.08.064] [Medline: 29169545]
15. Homer N, Yoon MK. Evaluation of the American Society of Ophthalmic Plastic and Reconstructive Surgery (ASOPRS) fellowship program website content and quality. Ophthalmic Plast Reconstr Surg 2017;33(6):471-473. [doi: 10.1097/IOP.0000000000000830] [Medline: 27997464]
16. Svider PF, Gupta A, Johnson AP, Zuliani G, Shkoukani MA, Eloy JA, et al. Evaluation of otolaryngology residency program websites. JAMA Otolaryngol Head Neck Surg 2014 Oct 01;140(10):956. [doi: 10.1001/jamaoto.2014.1714]
17. Skovrlj B, Silvestre J, Ibeh C, Abbatematteo JM, Mocco J. Neurosurgery residency websites: a critical evaluation. World Neurosurg 2015 Sep;84(3):727-733. [doi: 10.1016/j.wneu.2015.04.051] [Medline: 25940210]
18. Yayac M, Javandal M, Mulcahey MK. Accredited Orthopaedic Sports Medicine fellowship websites. Orthopaedic Journal of Sports Medicine 2017 Jan 27;5(1):232596711668394. [doi: 10.1177/2325967116683942]
19. Ashack KA, Burton KA, Soh JM, Lanoue J, Boyd AH, Milford EE, et al. Evaluating dermatology residency program websites. Dermatol Online J 2016 Mar 16;22(3):13030/qt7rx3j2dn. [Medline: 27136628]
20. Novin S, Yi PH, Vanderplas T, Yim D, Hong K. Integrated interventional radiology residency program websites: a development in progress. American Journal of Roentgenology 2018 Jul;211(1):211-216. [doi: 10.2214/ajr.17.19008]
21. Reilly EF, Leibrandt TJ, Zonno AJ, Simpson MC, Morris JB. General surgery residency program websites: usefulness and usability for resident applicants. Curr Surg 2004 Mar;61(2):236-240. [doi: 10.1016/j.cursur.2003.10.006] [Medline: 15051272]
22. Hu J, Zhen M, Olteanu C, Avram R. An evaluation of accessibility and content of microsurgery fellowship websites. Plast Surg (Oakv) 2016 Aug 01;24(3):187-190. [doi: 10.1177/229255031602400301]
23. Hansberry DR, Bornstein J, Agarwal N, McClure KE, Deshmukh SP, Long S. An assessment of radiology residency program websites. Journal of the American College of Radiology 2018 Apr;15(4):663-666. [doi: 10.1016/j.jacr.2017.11.010]
24. Oladeji LO, Yu JC, Oladeji AK, Ponce BA. How useful are orthopedic surgery residency web pages? J Surg Educ 2015 Nov;72(6):1185-1189. [doi: 10.1016/j.jsurg.2015.05.012] [Medline: 26089163]
25. Hashmi A, Policherla R, Campbell H, Khan FA, Schumaier A, Al-Mufarrej F. How informative are the plastic surgery residency websites to prospective applicants? J Surg Educ 2017 Jan;74(1):74-78. [doi: 10.1016/j.jsurg.2016.08.002] [Medline: 27717704]
26. Ward M, Pingree C, Laury AM, Bowe SN. Applicant perspectives on the otolaryngology residency application process. JAMA Otolaryngol Head Neck Surg 2017 Aug 01;143(8):782-787 [FREE Full text] [doi: 10.1001/jamaoto.2017.0231] [Medline: 28542682]
27. Shneiderman B. The Paradigm Birth of HCI and User Experience Design: Starting a Discipline and Launching an Industry. Medium. 2017 Apr. URL: https://tinyurl.com/tatp6b4t [accessed 2021-11-19]
28. Pang PC, Harrop M, Verspoor K, Pearce J, Chang S. What are health website visitors doing: insights from visualisations towards exploratory search. In: OzCHI '16: Proceedings of the 28th Australian Conference on Computer-Human Interaction.




XSL•FO
RenderX

2016 Presented at: 28th Australian Conference on Computer-Human Interaction; November 29-December 2, 2016; Launceston, Tasmania, Australia p. 631-633.

## Abbreviations

**GME:** graduate medical education
**HCI:** human-computer interaction
**UAMS:** University of Arkansas for Medical Sciences
**WCAG:** Web Content Accessibility Guidelines

*Edited by G Eysenbach; submitted 19.04.21; peer-reviewed by S Jalil, PCI Pang; comments to author 07.07.21; revised version received 17.08.21; accepted 18.09.21; published 13.12.21.*

*Please cite as:*
*Massanelli J, Sexton KW, Lesher CT, Jensen HK, Kimbrough MK, Privratsky A, Taylor JR, Bhavaraju A*
*Integration of Web Analytics Into Graduate Medical Education: Usability Study*
*JMIR Form Res 2021;5(12):e29748*
*URL: https://formative.jmir.org/2021/12/e29748*
*doi:10.2196/29748*
*PMID:34898459*

©Jackson Massanelli, Kevin W Sexton, Chris T Lesher, Hanna K Jensen, Mary K Kimbrough, Anna Privratsky, John R Taylor, Avi Bhavaraju. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Preventive Digital Mental Health for Children in Primary Schools: Acceptability and Feasibility Study

Sian M Davies[1,2], BSc, MSc; Jenni Jardine[2], DClinPsy; Kerry Gutridge[1,2], BSc, MA, PhD; Zara Bernard[1,2], BSc, MSc; Stephen Park[3]; Tom Dawson, PhD; Kathryn M Abel[1,2], BSc, MA, MB, MRCP, MRCPSYCH, PhD; Pauline Whelan[2,4], BA, MSc, PhD

[1]GM.Digital Research Unit, Centre for Women's Mental Health, Division of Psychology and Mental Health, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

[2]Greater Manchester Mental Health NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom

[3]Lexplore Ltd, Marple, United Kingdom

[4]GM.Digital Research Unit, Centre for Health Informatics, Division of Informatics, Imaging and Data Sciences, The University of Manchester, Manchester, United Kingdom

**Corresponding Author:**
Sian M Davies, BSc, MSc
GM.Digital Research Unit, Centre for Women's Mental Health
Division of Psychology and Mental Health, School of Health Sciences
Faculty of Biology, Medicine and Health, The University of Manchester
Jean McFarlane Building, Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 0161 306 7974
Email: sian.davies-5@manchester.ac.uk

## Abstract

**Background:** The incidence of mental health problems in children and adolescents in the United Kingdom has significantly increased in recent years, and more people are in contact with mental health services in Greater Manchester than in other parts of the country. Children and young people spend most of their time at school and with teachers. Therefore, schools and other educational settings may be ideal environments in which to identify those experiencing or those at the risk of developing psychological symptoms and provide timely support for children most at risk of mental health or related problems.

**Objective:** This study aims to test the feasibility of embedding a low-cost, scalable, and innovative digital mental health intervention in schools in the Greater Manchester area.

**Methods:** Two components of a 6-week digital intervention were implemented in a primary school in Greater Manchester: Lexplore, a reading assessment using eye-tracking technology to assess reading ability and detect early atypicality, and Lincus, a digital support and well-being monitoring platform.

**Results:** Of the 115 children approached, 34 (29.6%) consented and took part; of these 34 children, all 34 (100%) completed the baseline Lexplore assessment, and 30 (88%) completed the follow-up. In addition, most children were classified by Lincus as *regular* (≥1 per week) survey users. Overall, the teaching staff and children found both components of the digital intervention engaging, usable, feasible, and acceptable. Despite the widespread enthusiasm and recognition of the potential added value from staff, we met significant implementation barriers.

**Conclusions:** This study explored the acceptability and feasibility of a digital mental health intervention for schoolchildren. Further work is needed to evaluate the effectiveness of the digital intervention and to understand whether the assessment of reading atypicality using Lexplore can identify those who require additional help and whether they can also be supported by Lincus. This study provides high-quality pilot data and highlights the potential benefits of implementing digital assessment and mental health support tools in a primary school setting.

***(JMIR Form Res 2021;5(12):e30668)*** doi:10.2196/30668



XSL•FO
RenderX

**KEYWORDS**

digital mental health; acceptability; feasibility; child and adolescent mental health and well-being; school-based mental health care; prevention; digital assessment and monitoring; reading screening or ability

# Introduction

## Background

The prevalence of poor mental health in children and adolescents has significantly increased in the United Kingdom and, in recent years, has become a public health concern worldwide [1,2]. Greater Manchester has long been identified as an area of high unmet mental health need, with more people in contact with mental health services than in many other parts of the United Kingdom [3]. Common mental disorders in children and adolescents aged 5 to 14 years have been established as a leading cause of disability [4], with 50% of all adult mental health–related problems being diagnosed before the age of 14 years [5]. The rates of probable mental disorders in children and young people (CYP) have steadily and consistently risen, with 1 in 9 children aged 5 to 16 years being identified as having a probable mental disorder [6], which increased to 1 in 6 more recently [7].

The extensive literature has identified links between children's mental health, academic performance, and outcomes [8-11]. Many factors associated with poor mental health, such as social deprivation [12], are linked with poorer reading ability [13]. Similarly, there is an association between poorer reading ability and developmental and behavioral disorders that develop during childhood and adolescence, such as attention deficit hyperactivity disorder [14]. Therefore, early identification of reading difficulties may be instrumental in recognizing existing or developing mental health risks. In addition, it is widely accepted that early intervention and prevention strategies for CYP are critical: responding to early signs of distress can prevent symptoms from escalating and improve future outcomes [15]. Therefore, research into preventive mental health interventions for children and adolescents is becoming increasingly relevant and necessary [16].

Adolescents access primary care and other services for preventive health and well-being much less than other age groups [17]. As children and adolescents spend most of their day in school, schools are increasingly seen as ideal settings for delivering mental health support or interventions to young people [18-20]. The regular contact between teachers and students also means that school staff may identify those who are experiencing or at risk of developing mental health problems [21,22].

Given the low cost and scalability of digital interfaces and the widespread popularity of digital technologies among children and adolescents, digital interventions may provide new opportunities for delivering mental health support in schools [23]. Although digital support for children and adolescents has evolved and grown exponentially [24], there is a lack of research on how it can be safely and sustainably embedded in school settings. Furthermore, although the number of digital mental health apps for CYP is ever increasing, there still remains a gap in the evidence base behind them, despite their general acceptability [25], and few have been tested directly in school settings.

## Objective

This study aims to test the feasibility of embedding a low-cost, scalable, and innovative preventive digital mental health intervention in schools in the United Kingdom's Greater Manchester area. We piloted an innovative reading ability assessment tool (Lexplore) and a web-based well-being monitoring platform (Lincus) supported by the research team, with additional training for the school staff.

# Methods

To guide our analysis regarding the sustainability and adoption of the digital reading tool and the Lincus intervention, we performed a retrospective analysis using the nonadoption, abandonment, scale up, spread, and sustainability (NASSS) framework [26]. The NASSS framework has been widely used to examine the sustainability of digital health interventions in health care settings, and its applicability in schools has been discussed elsewhere [27,28].

## Digital Components

The digital intervention comprised two components: (1) Lexplore, a reading screening assessment that uses eye-tracking technology to assess reading ability, and (2) Lincus, a digital support and well-being monitoring platform.

### *Digital Assessment: Lexplore*

The Lexplore reading assessment uses eye-tracking technology to monitor a child's eye movements while they are reading. Eye movements can provide insight into the cognitive processes behind a child's individual reading method. Lexplore assesses reading ability (age- and sex-standardized) by calculating how long a child's eyes fixate on words and how they move through the passage. It can examine differences in how a child's brain processes text at lexical, syntactic, semantic, and structural levels. On the basis of this information and using machine learning, Lexplore can determine a child's reading attainment across 5 standardized levels ranging from low to high and can highlight with precision (and often before the child, teacher, or parent has noticed) the pupils who are experiencing reading difficulties. Lexplore supports the teachers' professional judgment and can also identify those individuals struggling with reading about whom the school or parents were unaware, often as children develop a set of coping strategies to manage their difficulties. Lexplore has recently been rolled out nationally across Swedish primary schools and is being increasingly used throughout the United Kingdom [29,30].

### *Digital Support: Lincus*

Lincus is a Conformitè Europëenne marked, class 1 medical device and health and social care management platform that measures well-being across 3 domains: emotional, social, and physical. It has been implemented in several health and social



XSL•FO
RenderX

care organizations, demonstrating utility in populations including individuals with long-term conditions, learning disabilities, homelessness, and multiple complex needs [31-35]. Previous research has demonstrated benefits such as improvements in reported mental health and general well-being, increased activity and perceived control of life, and better engagement and communication between health care professionals and service users [36]. Furthermore, Lincus is a configurable and customizable tool, and, specifically for this project, it was populated with child- and parent-relevant content (adapted to a simple child-friendly monitoring tool with animations for the sliding feedback scale).

## School Recruitment

The selection of schools used an opportunistic sampling method based on the networks associated with the research unit leading the study. A total of 2 schools were selected from our established links with the Manchester Healthy Schools program in Greater Manchester. Both schools were selected based on their awarded *gold status*, demonstrating their commitment to health promotion work [37].

## Participant Recruitment

All children were eligible for the study. Notably, we did not exclude children with low reading ability, as we aimed to build a digital framework that is accessible to all children. The Lincus platform is animated and image based to ensure that it is accessible to children who lack good reading skills. The research team visited the schools to provide an overview of the project to key senior staff. As they had all of the children's relevant contact details, the schools made the initial contact with the families and children by sending correspondence from the research team to the parents or guardians. The correspondence included a letter to the parent or guardian, parental or guardian information leaflet, child information leaflet, parental or guardian consent form, and an example child assent form. Parents or guardians were informed about the project by an invitation letter and information sheet for parents or guardians. The information sheet detailed the nature and objectives of the study and possible risks associated with participation. If parents or guardians were happy with all the arrangements and for themselves and their child to take part, they were then asked to complete the parental or guardian's informed consent form and return it to the schools.

## Ethics

The study received ethical approval from the University of Manchester Research Ethics Committee (2019-7489-11848).

## Procedure

The digital framework had two components: Lexplore initial assessment and Lincus.

### *Lexplore Initial Assessment*

Participants initially completed a short reading ability task using the Lexplore digital platform, which involved children reading text from a computer screen with an eye-tracking sensor attached to it. On average, it took approximately between 2 and 5 minutes to complete the task. The teachers were provided with training by the Lexplore staff to enable them to perform this assessment test. After 6 weeks, participants were asked to complete the Lexplore reading ability task again.

### *Lincus*

The Lincus platform was customized for child use and populated with Greater Manchester local health and well-being resources and links. Each participant was provided their own secure username, which corresponded to their participant ID number and password. This was also shared with their parents or guardians to enable them to have the option to use Lincus outside of school hours should they wish to. The Lincus platform was accessible via a web browser on a computer or tablet. Teachers were also provided with a secure username and password to have access to pseudonymized data on the platform and were provided with the appropriate training to be able to use Lincus.

Participants were asked to spend approximately 5 minutes during their free time in the morning each day for 6 weeks and complete 2 self-report surveys (lifestyle support and well-being) on the Lincus platform. Participants were encouraged to complete the surveys independently; however, the teachers were able to assist if required. The platform also included options to record observations and web-based support links for children, parents or guardians, and teachers.

Following the intervention and follow-up assessment, children, their parents or guardians, and their teachers were invited to a qualitative focus group or workshop to provide feedback on the digital framework.

## Data Collection

Data were collected via the Lexplore and Lincus platforms separately. The data from Lexplore included a score of reading age, ability, comprehension, and speed for each participant during the testing and follow-up phases of the project. The Lincus platform collected self-report data on emotional, social, and physical well-being. Lincus data also included usability and engagement metrics (ie, how many times the child used the platform, modules accessed, and duration of use). If a participant withdrew during the project, no further data were collected from them; however, historic data were retained. Qualitative feedback from parents was collected via their evening appointments rather than workshops, as the school advised this would be the most suitable way to collect parent feedback. Feedback from the teaching staff was also obtained during the parents' evening appointments because of the limited capacity and time for conducting a separate focus group or workshop. Feedback from children was gathered as part of a group workshop.

## Data Analysis

We used the NASSS framework [26] as a post hoc method to analyze the data and understand the barriers and facilitators to implementing the intervention (Textbox 1).



XSL•FO
RenderX

**Textbox 1.** The nonadoption, abandonment, scale up, spread, and sustainability framework.

**Condition**

- The prevalence of poor mental health in children and adolescents has significantly increased in the United Kingdom and, in recent years, has become a public health concern worldwide, with Greater Manchester being a particular area of high need.

**Technology**

- An integrated digital framework where both had good usability, and Lincus had been co-designed and tailored for the Greater Manchester population. Both were standalone systems outside the usual technologies used in schools.
- Data made available were reported as clear and helpful; however, engagement from teaching assistants and parents was poor. Therefore, education about how to access the data was found wanting.
- Minor issues were experienced with access to Lincus, such as log-in difficulties and problems with firewall settings. These difficulties were sometimes readily solved or would have been relatively easy to solve with appropriate communication from the staff.
- The technology was supplied through project grant funding; however, both were low-cost technologies. Lexplore has already been extended and deployed at other schools.

**Value proposition**

There was evidence of both demand-side and supply-side value:

- To secure funding for the project, a clear business case was presented for the Lincus system and Lexplore assessment tool (supply-side value).
- The desirability (demand-side value) existed, with enthusiasm from senior staff regarding the technologies being low cost and scalable, and innovative digital screening platforms to identify and support children who were most vulnerable to mental health problems.
- However, some parents did not perceive the digital system to have value, as they did not consider their children to be needing mental health support.

**Adopter system**

- The staff found the disruption of their morning routine as adding to pressures and demands.
- Teaching assistants are responsible for a large number of children and tasks and can perceive additional responsibilities as burdensome.

**The organization**

- Staff reported that existing pressures limited their ability to engage fully with the project.
- Recruitment and consent process was deemed time consuming.

**Wider system**

- Children and young people's mental health and well-being is and continues to be a key priority; Greater Manchester is, in particular, an area of high unmet need.
- Since the COVID-19 pandemic, digital technologies are increasingly being used to support children and young people within educational and health care settings.

**Embedding and adaptation over time**

- There is scope for adaptation over time of both the system and in the way the technology is deployed.
- This study found a lack of organizational resilience to changes and the embedding of new technologies because of staff capacity and workload.

# Results

## Overview

The project was aimed at children aged between 6 and 10 years. The first school completed testing in January 2020, with 34 children and 5 teachers across year 4 or 3rd grade and year 5 or 4th grade (mean age 9 years; 19/34, 56% females, 15/34, 44% males). The uptake was 29.6% (34/115; total number of children approached being 115 across the 4 classes). The second school agreed to participate from the year 4 or 3rd grade and year 5 or 4th grade, and their consent forms were completed. However, staffing difficulties and sicknesses meant they had to withdraw.

A total of 4 more schools were contacted across Greater Manchester. A total of 2 schools did not respond and 2 schools expressed interest; however, 1 could not confirm participation because they were waiting for a new head teacher to commence. The other later declined, citing that the school was extremely busy with other demands. Access to further schools became impossible following the COVID-19 lockdown in the United Kingdom from March 23, 2020.

## User Engagement

During the 6-week period of data collection, Lexplore assessments were conducted twice in each school. All children completed an initial assessment, and 88% (30/34) completed



XSL•FO
RenderX

their follow-up reading assessments. The follow-up consent forms were not returned by parents for the second round of assessment; these children were not tested at follow-up. A total of 322 Lincus well-being surveys were recorded. On average, 46 well-being surveys were recorded per week, equating to approximately 1 well-being survey per user per week in total (6 per user per month). Most children were classified by Lincus as *regular* (≥1 per week) survey users. There was a significant drop in use over the 2-week Christmas holidays, although 6% (2/34) of children reported logging in at home; this can explain the lower average during December.

Approximately 18% (4/34) of children did not complete any well-being surveys; the reasons provided were that they had never logged in because of late attendance, sickness, or not having time in the morning to complete the survey tasks.

## Poststudy Consultations With the Adopters (the Adopter System—NASSS Framework)

### *Children*

Approximately 68% (23/34) of children participated in group feedback (14/23, 61% from year 4 or 3rd grade and 9/34, 39% from year 5 or 4th grade). Overall, there was a favorable opinion among the children regarding the digital platforms, as indicated by their quotes (Textbox 2).

**Textbox 2.** Quotes from children describing how they felt about Lexplore, Lincus, and confidentiality.

- "Reading was fun. It was futuristic"; "It picked up I have a lazy eye"; "it's cool!"
- "It felt good and weird to tell the internet how I feel"; "You can tell the truth about yourself instead of telling the whole world"; "Sometimes I don't want to say how I'm feeling so instead of talking, I liked putting down and people can read it."
- "I didn't answer the truth on sleep in case I got into trouble for staying up and being tired in the classroom the next day so said I slept better."
- "I lied about my social life and put the score where I wanted it to be and not where it was."
- "Anxiety—I put it lower I was embarrassed, I didn't want to score it in case anyone found out, my friend had anxiety and they got bullied so didn't want the same to happen to me."

Qualitative feedback demonstrated that the Lincus content was mostly found acceptable and relevant by users, although some domains of well-being needed explanation by the teaching assistant (TA; eg, control of life and appetite). Children reported that the pictures helped them understand the scale. Completion time of the measures was found to be acceptable for children reporting quick response times. They described finding it easy to complete the scales, and once they were able to access the platform following log-in, they required no further support.

The request for daily log-ins to Lincus was deemed burdensome by some children, with suggestions of a couple of times a week being preferable. Log-in difficulties were highlighted, resulting in children being locked out of their accounts and, consequently, being unable to complete the survey during their allocated time. In addition, some children reported that having to engage with the platforms meant that they missed out on other enjoyable activities alongside their nonparticipating peers.

Some children expressed concerns regarding the confidentiality of their Lincus data; specifically, they were worried that their teachers, parents, and other third parties might have access to the data, which may have influenced their responses. Moreover, some children suggested that they inflated their scores to be more desirable (Textbox 2). The duration of the Lexplore assessment was approximately 5 minutes; several children commented that they would have liked the process to be even quicker. However, 2 sessions were acceptable to them.

### *Parents*

The school suggested that we gather feedback at the parents' evening appointments to maximize parental engagement. The parents of 21% (7/34) of children were interviewed regarding the study. None of these parents had accessed their child's well-being survey, citing reasons such as being too busy, having no concern regarding their child's well-being, and log-in difficulties. Furthermore, parents reported that they were not aware of the health and well-being resources held on Lincus. At the time of the interview, parents had not yet received any feedback from teachers regarding their child's Lexplore results, despite this being part of the teachers' training by research staff at the beginning of the program.

### *Teachers*

Out of 4 teachers across the year groups, 2 (50%) provided feedback at the parents' evening appointment (one year 4 or 3rd grade and one year 5 or 4th grade teacher). The year 4 or 3rd grade teacher rated both platforms positively overall. However, they had forgotten that Lincus hosted a wealth of mental health and well-being resources despite receiving training on the platform. The staff were provided with personal log-in details; however, none of them had accessed the platform. The year 5 or 4th grade teacher felt less informed, as they did not attend the training session and, therefore, were not aware of the potential value and capabilities of the Lincus platform. They suggested that rather than having to log on to each child's profile regularly, which was considered by some as too time consuming, receiving notifications regarding a child who was scoring consistently low would be more beneficial.

The Lexplore results were reported to be consistent with their own assessment of the children's reading performance; however, they liked that it added an objective measure. They believed these scores could be helpful in following up on children who were struggling and aid in discussion with their parents. The teachers reflected that if everyone in the class had opted in, the completion of the assessments would have been less disruptive, as they would have been able to complete all assessments together in 1 classroom. However, they also felt that it would



XSL•FO
RenderX

not have been feasible because of the limited staff capacity to monitor and support all the children completing the assessment.

### *Teaching Staff*

Both TAs were solely responsible for supporting the children in using the Lincus survey and completing their reading assessments. Feedback was based on 1 TA who worked across the two year 4 or 3rd grade classes and was available at the time of the consultation. This consultation identified the implementation strategies, barriers, and facilitators.

### Implementation Strategies

TAs were asked to support the daily input of Lincus well-being data as per the study protocol. However, as the study progressed, the TAs reported that the children became more self-directed and required fewer reminders. They were able to seek out the tablets independently. All well-being surveys were conducted before formal lessons. Additional training was provided for the TAs to complete the Lexplore reading assessments at the 2 assessment points. Each class was allocated an assessment period, which took approximately half a day to complete.

### Implementation Barriers and Facilitators

Log-in difficulties were a barrier that was identified early on in the intervention, which locked some children out of completing their survey in the time allocated. Log-in information or passwords were set by the system and were written down and accessible to the children; however, because the password entry box was blinded to them, this increased the frequency of mistakes. The suggestions were that children could choose their own password or log-in information that they would be able to remember and easily type out to mitigate this. Once the children had completed a few surveys, it was evident that they could input their well-being data without further support.

Teachers reported that they found it challenging to incorporate daily well-being surveys into their demanding morning routines, although the staff did not feel that other times of the day were more suitable. Completing Lexplore was not deemed time consuming for the actual assessment. However, the organizational logistics around implementing assessments were viewed as labor intensive (eg, finding a suitable and available room, collecting and preparing children for the assessment, and then returning them to classrooms). This was perceived negatively, as typical reading assessments are spaced out and do not require equipment or extra organization.

The availability of digital equipment (tablets) across classes at the same time was highlighted as a challenge. Log-in difficulties that timed the children out of the well-being surveys reduced user engagement. Poor parental engagement in providing consent was an additional barrier in the recruitment context, resulting in a delayed start of the study. This meant that there was a 3-month interval between staff training on the digital platforms and the study's commencement, leading to the staff feeling less familiar with the platform and its operation.

No difficulties were explicitly reported in maintaining the children's engagement as a barrier to implementation. Parental involvement was not a necessity for the project once it had started, clearly demonstrated by the high user engagement of children in the school setting.

## Discussion

### Principal Findings

#### *Overview*

The primary aim of this project was to examine the feasibility and acceptability of embedding a low-cost, scalable, and innovative assessment and preventive digital mental health tool for schoolchildren in the Greater Manchester area. Overall, children found this digital intervention engaging, usable, and acceptable. However, despite widespread enthusiasm and recognition of the potential added value from head teachers during the consultation phases of the project, we met significant implementation barriers. Consistent with the findings of Edridge et al [27,28], all 6 themes representing implementation barriers within the NASSS framework emerged: technology, value proposition, the adopter system, the organization, wider system, and embedding and adaptation over time.

#### *Technology*

Minor issues were experienced with log-in access to Lincus. These difficulties were sometimes readily solved or would have been relatively easy to solve with appropriate communication from the staff. Lack of communication by the staff resulted in some difficulties remaining unresolved. This then became a more significant barrier and prevented or reduced user engagement. Difficulties in accessing Lincus could be readily mitigated by providing personal log-in details. Additional barriers were identified by the second school (which later withdrew), with a security firewall preventing access to the platform. This is a straightforward problem to resolve from a technical perspective by enabling access to the platform through the firewall. However, it presented a significant implementation barrier, as the teaching staff did not readily know how to identify, report, or resolve the access problem. These are important considerations for future work across educational settings, as such barriers may feel overwhelming or burdensome to staff already under significant pressure and may influence their participation.

#### *Value Proposition*

There was evidence of both demand-side and supply-side value. To secure funding for the project, a clear business case was presented for the Lincus system and the Lexplore assessment tool (supply-side value). The desirability (demand-side value) existed, with enthusiasm from senior staff regarding the technologies representing low-cost, scalable, and innovative digital screening platforms for the early identification of and support for children most vulnerable to mental health problems. However, the benefits could have been reinforced with the TA staff to improve motivation to engage with the project. In addition, some parents did not perceive the digital system as having value as they did not believe their children needed mental health support.



XSL•FO
RenderX

### *The Adopter System*

Overall, children reported finding both Lexplore and Lincus acceptable, and their feedback suggested that they would continue to engage with both platforms. However, barriers were identified in engagement for the staff and parents. For instance, participation required some change to the morning routine for teaching staff, which added strain to an already demanding schedule. TAs bear the weight of many responsibilities for a large number of children and, therefore, may be at risk of perceiving additional tasks as burdensome. In our study, we found that TA availability and willingness are integral to the feasibility of school-based mental health programs. Therefore, recruiting TAs who support the aims and have a clear understanding of the program benefits is key to success. Ensuring that TAs have sufficient time to support children among their other daily tasks will also be an important element of future implementation. Enthusiasm from head teachers would have to be met by their leadership in championing the assessment tools and supporting the teachers and TAs in delivering the program. Although parents could be categorized as being part of the adopter system, their role was not required for children's successful engagement with the platforms. Feedback from the interviewed parents was that they were not concerned regarding their children's mental health and well-being. This could explain why they did not access the platforms themselves, as they did not consider their children to have any mental health–related concerns and, therefore, felt there was no need to check up on them. As outlined above, previous work has highlighted the concerns of parents and schools regarding children and adolescent mental health. However, the sample we were able to recruit for this feasibility pilot may not be representative of the parents we originally aimed to target, as this study's parents did not report having worries or concerns regarding their children. An explanation might be that families from more socioeconomically deprived or ethnically diverse backgrounds are less likely to take part in research so that children who may be more likely to experience mental health problems are less likely to engage in research and are underrepresented in the samples [38]. Future research should consider the barriers to participation from underserved groups; this is increasingly important to ensure that samples are inclusive and involve those with unmet needs and who may be most likely to benefit from these interventions.

### *The Organization*

The staff reported that significant pressures on them, though unrelated to this study, limited their ability to engage fully with the project. The consenting and recruitment process for children was time consuming and required multiple periods of engagement with the school staff, resulting in relatively low rates of consent. Successful roll out of digital platforms for future routine use in schools is likely to need more extensive engagement with teachers and TAs. In our view, the organizational barriers we identified are largely surmountable, some more readily than others. We recommend future research projects to ensure adequate upfront engagement with schools so that all school staff, including senior leadership, frontline teaching staff, and TAs, are fully supported to deliver the research.

### *Wider System*

The wider context was and continues to be supportive of mental health and well-being within schools. Particularly since the COVID-19 pandemic, digital technologies are increasingly being deployed to support CYP within schools and across mental health services [25]. CYP's mental health and well-being are key priorities and continue to be so. Greater Manchester is a particular area of high unmet need for CYP mental health [3]. Not feeling listened to and perceived social stigma have previously been identified as key barriers to CYPs engaging with specialist services and seeking help [39]. Personalized digital tools such as Lexplore and Lincus allow CYP to have their voices heard in a safe and nonstigmatized way, in contrast to meeting school nurses or counselors or their general practitioners with their parents.

### *Embedding and Adaptation Over Time*

There is scope for adaptation over time of both the systems and in the way the technology is deployed. However, we found a lack of organizational resilience regarding changes and the embedding of new technologies in this study because of limited staff capacity and a focus on other school priorities.

## Strengths, Limitations, and Recommendations

There are several strengths to this feasibility study. We used an established digital health intervention framework to evaluate the adoption, scale up, spread, and sustainability, and in the school where engagement was possible, we recruited nearly one-third of the eligible children in a relatively short space of time. The Lincus platform was easily adaptable to embed local information that was relevant and contextualized to schools and children. Support from school leadership highlighted the need for such digital programs to supplement ongoing mental health delivery in schools. Preventive strategies are key in early identification to provide timely support to reduce the risk of development or escalation of mental health problems. Both the Lexplore and Lincus interventions may be beneficial in enabling the staff and parents to identify issues by monitoring reading ability and well-being. The feasibility study demonstrated the acceptability of the digital intervention to staff and children and the willingness of children and parents to consent to engage with digital tools.

This study also has some important limitations. We do not have information regarding children who did not participate but were eligible to do so. This makes the assessment of bias and representativeness of the sample unclear. Demographic data, such as ethnicity and socioeconomic status were not captured. The overall recruitment of children was limited because of the challenges in engaging key school staff and parents. Furthermore, the intervention was only tested in 1 region of the country with opportunistic sampling and, therefore, does not indicate how this would work in other settings across the United Kingdom. Although the qualitative data show a positive appraisal of both platforms, and children generally had a positive experience using the tools, this may not have been the case for all children who were eligible to participate. Similarly, we had limited participation from staff and parents, making it difficult to determine the acceptability of the digital tools for parents,



XSL•FO
RenderX

carers, and teachers. This may not be a problem in the future if children accept and adopt the tools and parents and teachers follow their lead. Another key consideration is funding and resource availability: the success of both implementing and adopting digital technologies in a school setting means all children can access the equipment they require to participate, such as tablets, laptops, and computers. This might mean additional support and resources for children from lower socioeconomic backgrounds. We conducted this as a *standalone* program; it may work better when integrated as part of personal, social, health, and economic education or other extant modules within the school curriculum. Finally, it was not specifically linked to mental health programs in schools; this is an important consideration if the platforms are to become successfully embedded in schools. Using the platform as an educational tool or integrating it into other digital educational tools as part of the curriculum would guarantee adoption with added health benefits. Embedding digital health interventions within the existing school structure and programs may be a better way to increase involvement and commitment from staff, parents, and children.

### Future Work

The daily demands faced by the teaching staff are key barriers to embedding any new technology requiring significant teacher input in schools. Despite the senior staff's willingness to welcome the new technology and a desire from some teachers and TAs to take part in the research (which they saw as relevant to their practice), much more work is needed to demonstrate the value of digital platforms to staff and parents. Future deployments should work closely with schools, children, and parents from the outset to codevelop the platform with them to respond to their needs and facilitate adaptation over time, as opposed to implementing ready-made tools. Identifying research champions in classes or year groups could offer a route to more seamless engagement [40]. This suggests that future studies could bypass the need for ongoing TA involvement, and this may also be likely to encourage participation from other children.

The COVID-19 pandemic has created far greater digital engagement from schools, teachers, pupils, and parents. This may present an opportunity for teaching staff to be able to engage more readily in the future with web-based platforms such as Lincus and Lexplore.

## Conclusions

The key aim of this study was to embed a low-cost, scalable, and innovative digital mental health intervention in schools in the Greater Manchester area to identify and provide timely support for children most at risk of mental health problems. Overall, the digital platforms were well-received, and the study revealed important barriers and facilitators that can provide key information and associated recommendations for conducting future research in this setting. The landscape has changed dramatically during the COVID-19 pandemic, with a spike in interest in the use of digital technologies to manage health and well-being. Some of the difficulties we encountered in the feasibility of widening the implementation of digital mental health and educational support tools may have now been mitigated. Where staff felt adequately supported, both platforms could be delivered feasibly, and overall, children and parents found them acceptable. However, the teaching staff play an instrumental role in the success of implementing digital technologies, and staff attitudes influence the degree to which new technologies are accepted within traditional working practices. Therefore, future options should minimize the need for staff and parent involvement and focus on widening children's participation. Furthermore, staff training, funding, resources, and staff willingness to engage and participate must all be considered for the successful implementation of digital mental health solutions in schools. Although this study did encounter some difficulties, it provided interesting pilot data that highlight the potential benefits of implementing digital health tools within a school setting.

## Acknowledgments

The authors would like to thank the OneMedical Group and all the teaching staff, parents, and children who were involved in this research. This project has received funding from Health Innovation Manchester and support for the recruitment of children from the Greater Manchester Clinical Research Network.

## Conflicts of Interest

PW is a director of Affigo CIC and Careloop Ltd, digital mental health companies developed out of the University of Manchester. PW is also a director of Prism Life Ltd, a small consultancy company. Affigo, Careloop, and Prism had no role in this project.

## References

1. Williams I, Vaisey A, Patton G, Sanci L. The effectiveness, feasibility and scalability of the school platform in adolescent mental healthcare. Curr Opin Psychiatry 2020 Jul;33(4):391-396. [doi: 10.1097/YCO.0000000000000619] [Medline: 32398545]
2. Polanczyk GV. Addressing the evidence gap on preventive interventions. J Am Acad Child Adolesc Psychiatry 2015 Oct;54(10):793-794. [doi: 10.1016/j.jaac.2015.07.014] [Medline: 26407486]
3. Greater Manchester mental health research priorities listening exercise: considering priorities for mental health research for the Greater Manchester population. Health Innovation Manchester. 2019. URL: https://healthinnovationmanchester.com/wp-content/uploads/2019/09/Mental-Health-Research-Listening-Exercise-Sept-2019.pdf [accessed 2021-08-20]



XSL•FO
RenderX

4. Baranne ML, Falissard B. Global burden of mental disorders among children aged 5-14 years. Child Adolesc Psychiatry Ment Health 2018 Apr 12;12(1):19 [FREE Full text] [doi: 10.1186/s13034-018-0225-4] [Medline: 29682005]
5. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
6. Mental health of children and young people in England, 2017 [PAS]. NHS Digital. 2018. URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017#summary [accessed 2021-08-16]
7. Mental health of children and young people in England, 2020: wave 1 follow up to the 2017. NHS Digital. 2020. URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2020-wave-1-follow-up [accessed 2021-08-16]
8. Frank JL, Kohler K, Peal A, Bose B. Effectiveness of a school-based yoga program on adolescent mental health and school performance: findings from a randomized controlled trial. Mindfulness 2016 Oct 25;8(3):544-553. [doi: 10.1007/s12671-016-0628-3]
9. Sznitman SR, Reisel L, Romer D. The neglected role of adolescent emotional well-being in national educational achievement: bridging the gap between education and mental health policies. J Adolesc Health 2011 Feb;48(2):135-142. [doi: 10.1016/j.jadohealth.2010.06.013] [Medline: 21257111]
10. Brännlund A, Strandh M, Nilsson K. Mental-health and educational achievement: the link between poor mental-health and upper secondary school completion and grades. J Ment Health 2017 Aug 07;26(4):318-325. [doi: 10.1080/09638237.2017.1294739] [Medline: 28266232]
11. Knopf JA, Finnie RK, Peng Y, Hahn RA, Truman BI, Vernon-Smiley M, Community Preventive Services Task Force. School-based health centers to advance health equity: a community guide systematic review. Am J Prev Med 2016 Jul;51(1):114-126 [FREE Full text] [doi: 10.1016/j.amepre.2016.01.009] [Medline: 27320215]
12. Dolean D, Melby-Lervåg M, Tincas I, Damsa C, Lervåg A. Achievement gap: socioeconomic status affects reading development beyond language and cognition in children facing poverty. Learn Instr 2019 Oct;63:101218. [doi: 10.1016/j.learninstruc.2019.101218]
13. Boyes ME, Tebbutt B, Preece KA, Badcock NA. Relationships between reading ability and child mental health: moderating effects of self‐esteem. Aust Psychol 2020 Nov 12;53(2):125-133. [doi: 10.1111/ap.12281]
14. Hart SA, Petrill SA, Willcutt E, Thompson LA, Schatschneider C, Deater-Deckard K, et al. Exploring how symptoms of attention-deficit/hyperactivity disorder are related to reading and mathematics performance: general genes, general environments. Psychol Sci 2010 Nov 21;21(11):1708-1715 [FREE Full text] [doi: 10.1177/0956797610386617] [Medline: 20966487]
15. Universal approaches to improving children and young people's mental health and well-being lay : summary report of the synthesis of systematic reviews and grey literature review. Public Health England. 2019. URL: https://dera.ioe.ac.uk/34484/1/Lay_summary_report.pdf [accessed 2021-05-17]
16. Fjermestad KW, Wergeland GJ, Rogde A, Bjaastad JF, Heiervang E, Haugland BSM. School-based targeted prevention compared to specialist mental health treatment for youth anxiety. Child Adolesc Ment Health 2020 May 12;25(2):102-109. [doi: 10.1111/camh.12366] [Medline: 32307836]
17. Patton G, Sawyer S, Santelli J, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016 Jun 11;387(10036):2423-2478 [FREE Full text] [doi: 10.1016/S0140-6736(16)00579-1] [Medline: 27174304]
18. Fazel M, Hoagwood K, Stephan S, Ford T. Mental health interventions in schools in high-income countries. Lancet Psychiatry 2014 Oct;1(5):377-387. [doi: 10.1016/s2215-0366(14)70312-8]
19. Fazel M, Patel V, Thomas S, Tol W. Mental health interventions in schools in low-income and middle-income countries. Lancet Psychiatry 2014 Oct;1(5):388-398. [doi: 10.1016/s2215-0366(14)70357-8]
20. Schulte-Körne G. Mental health problems in a school setting in children and adolescents. Dtsch Arztebl Int 2016 Mar 18;113(11):183-190 [FREE Full text] [doi: 10.3238/arztebl.2016.0183] [Medline: 27118666]
21. Werner-Seidler A, Perry Y, Calear AL, Newby JM, Christensen H. School-based depression and anxiety prevention programs for young people: a systematic review and meta-analysis. Clin Psychol Rev 2017 Feb;51:30-47 [FREE Full text] [doi: 10.1016/j.cpr.2016.10.005] [Medline: 27821267]
22. Weisz JR, Sandler IN, Durlak JA, Anton BS. Promoting and protecting youth mental health through evidence-based prevention and treatment. Am Psychol 2005 Sep;60(6):628-648. [doi: 10.1037/0003-066x.60.6.628]
23. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]
24. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr 10;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]




XSL•FO
RenderX

25. Bergin AD, Vallejos EP, Davies EB, Daley D, Ford T, Harold G, et al. Preventive digital mental health interventions for children and young people: a review of the design and reporting of research. NPJ Digit Med 2020 Oct 15;3(1):133 [FREE Full text] [doi: 10.1038/s41746-020-00339-7] [Medline: 33083568]
26. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
27. Edridge C, Deighton J, Wolpert M, Edbrooke-Childs J. The implementation of an mHealth intervention (ReZone) for the self-management of overwhelming feelings among young people. JMIR Form Res 2019 May 02;3(2):e11958 [FREE Full text] [doi: 10.2196/11958] [Medline: 31045499]
28. Edridge C, Wolpert M, Deighton J, Edbrooke-Childs J. An mHealth intervention (ReZone) to help young people self-manage overwhelming feelings: cluster-randomized controlled trial. J Med Internet Res 2020 Jul 27;22(7):e14223 [FREE Full text] [doi: 10.2196/14223] [Medline: 32716299]
29. Benfatto MN, Seimyr GO, Ygge J, Pansell T, Rydberg A, Jacobson C. Screening for dyslexia using eye tracking during reading. PLoS One 2016 Dec 9;11(12):e0165508 [FREE Full text] [doi: 10.1371/journal.pone.0165508] [Medline: 27936148]
30. Case studies. Lexplore Analytics. 2021. URL: https://www.lexplore.com/gb/cases/ [accessed 2021-05-17]
31. Blanchard A, Prior F, Gilbert L, Dawson T. Technology implementation case studies: Lincus Software as a service. In: Dastbaz M, Arabnia H, Akhgar B, editors. Technology for Smart Futures. Cham: Springer; 2018:141-167.
32. Blanchard A, Gilbert L, Dawson T. UK HealthSocial Care case studies: iterative technology development. In: Randell R, Cornet R, McCowan C, Peek N, Scott P, editors. Informatics for Health: Connected Citizen-Led Wellness and Population Health (Studies in Health Technology and Informatics). Amsterdam: IOS Press; 2017:83.
33. Prior F, Dawson T. Development of a holistic health economic evaluation tool leveraging patient self-report. In: Proceedings of the 9th International Conference on Developments in eSystems Engineering (DeSE). 2016 Presented at: 9th International Conference on Developments in eSystems Engineering (DeSE); Aug. 31 - Sept. 2, 2016; Liverpool, UK. [doi: 10.1109/dese.2016.12]
34. Blanchard A, Prior F, Barton J, Dawson T. Barriers in new health and social care technology implementation in the United Kingdom, a case study. In: Proceedings of the 9th International Conference on Developments in eSystems Engineering (DeSE). 2016 Presented at: 9th International Conference on Developments in eSystems Engineering (DeSE); Aug. 31 - Sept. 2, 2016; Liverpool, UK. [doi: 10.1109/dese.2016.13]
35. Dawson TA. Symptom and lifestyle tracking. Visualization Network-of-Experts 11th Workshop: Visual Analytics and Network Operations and Health. 2011. URL: https://lincus.onemedicaltechnologies.co.uk/content/docs/techpapers/Visual_Analytics_and_Network_Operations_and_Health.pdf [accessed 2021-11-13]
36. Gilbert L, Blanchard A, Barton J, Dawson T. Lincus: an adaptation and pilot usability study on individuals with learning disabilities. Rescon Internal Publication. 2014. URL: https://lincus.onemedicaltechnologies.co.uk/content/docs/case/LincusHFT.pdf [accessed 2021-11-13]
37. Healthy schools. Manchester Healthy Schools. 2021. URL: https://www.manchesterhealthyschools.nhs.uk/achievements [accessed 2021-04-26]
38. Jang M, Vorderstrasse A. Socioeconomic status and racial or ethnic differences in participation: web-based survey. JMIR Res Protoc 2019 Apr 10;8(4):e11865 [FREE Full text] [doi: 10.2196/11865] [Medline: 30969173]
39. Radez J, Reardon T, Creswell C, Orchard F, Waite P. Adolescents' perceived barriers and facilitators to seeking and accessing professional help for anxiety and depressive disorders: a qualitative interview study. Eur Child Adolesc Psychiatry 2021 Jan 27:A. [doi: 10.1007/s00787-020-01707-0] [Medline: 33502596]
40. Burn K, Conway R, Edwards A, Harries E. The role of school‐based research champions in a school–university partnership. Br Educ Res J 2020 Aug 06;47(3):616-633. [doi: 10.1002/berj.3675]

## Abbreviations

**CYP:** children and young people
**NASSS:** nonadoption, abandonment, scale up, spread, and sustainability
**TA:** teaching assistant



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 24.05.21; peer-reviewed by A Bergin, L Balcombe, R Pine; comments to author 28.06.21; revised version received 23.08.21; accepted 13.09.21; published 13.12.21.*

*Please cite as:*
*Davies SM, Jardine J, Gutridge K, Bernard Z, Park S, Dawson T, Abel KM, Whelan P*
*Preventive Digital Mental Health for Children in Primary Schools: Acceptability and Feasibility Study*
*JMIR Form Res 2021;5(12):e30668*
*URL: https://formative.jmir.org/2021/12/e30668*
*doi:10.2196/30668*
*PMID:34898446*

©Sian M Davies, Jenni Jardine, Kerry Gutridge, Zara Bernard, Stephen Park, Tom Dawson, Kathryn M Abel, Pauline Whelan. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# The Life Goals Self-Management Mobile App for Bipolar Disorder: Consumer Feasibility, Usability, and Acceptability Study

Kelly A Ryan[1*], PhD; Shawna N Smith[2*], PhD; Anastasia K Yocum[1*], PhD; Isabel Carley[1*], BS; Celeste Liebrecht[3,4*], MSW; Bethany Navis[1*], BS; Erica Vest[1*], MSW; Holli Bertram[1*], MSW; Melvin G McInnis[1*], MD; Amy M Kilbourne[3,4*], PhD

[1]Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States

[2]Department of Health Management and Policy, School of Public Health, University of Michigan, Ann Arbor, MI, United States

[3]VA Ann Arbor Healthcare System, United States Department of Veterans Affairs, Ann Arbor, MI, United States

[4]Department of Learning Health Sciences, University of Michigan, Ann Arbor, MI, United States

[*]all authors contributed equally

**Corresponding Author:**
Kelly A Ryan, PhD
Department of Psychiatry
University of Michigan
2101 Commonwealth Blvd, Suite C
Ann Arbor, MI, 48105
United States
Phone: 1 734 936 5524
Fax: 1 734 763 9259
Email: karyan@umich.edu

## Abstract

**Background:** Life Goals is an evidence-based self-management intervention that assists individuals with bipolar disorder (BD) by aligning BD symptom coping strategies with their personal goals. The intervention can be availed via in-person and telephonic sessions, and it has been recently developed as an individualized, customizable mobile app.

**Objective:** We examined the feasibility, usability, and acceptability of the Life Goals self-management app among individuals diagnosed with BD who used the app for up to 6 months.

**Methods:** A total of 28 individuals with BD used the Life Goals app on their personal smartphone for 6 months. They completed key clinical outcome measurements of functioning, disability, and psychiatric symptoms at baseline, 3 months, and 6 months, in addition to a poststudy survey about usability and satisfaction.

**Results:** Participants used the app for a median of 25 times (IQR 13-65.75), and for a longer time during the first 3 months of the study. The modules on depression and anxiety were the most frequently used, accounting for 35% and 22% of total usage, respectively. Overall, the study participants found the app useful (15/25, 60%) and easy to use (18/25, 72%), and they reported that the screen displayed the material adequately (22/25, 88%). However, less than half of the participants found the app helpful in managing their health (10/25, 40%) or in making progress on their wellness goals (9/25, 36%). Clinical outcomes showed a trend for improvements in mental and physical health and mania-related well-being.

**Conclusions:** The Life Goals app showed feasibility of use among individuals with BD. Higher user engagement was observed in the initial 3 months with users interested more frequently in the mood modules than other wellness modules. Participants reported acceptability with the ease of app use and satisfaction with the app user interface, but the app showed low success in encouraging self-management within this small sample. The Life Goals app is a mobile health technology that can provide individuals with serious mental illness with more flexible access to evidence-based treatments.

***(JMIR Form Res 2021;5(12):e32450)*** doi:10.2196/32450

**KEYWORDS**

self-management; app; bipolar disorder; symptom management; mental health; feasibility; usability; acceptability; intervention; bipolar; coping; survey; engagement



XSL•FO
RenderX

## Introduction

Bipolar disorder (BD), a serious mental illness, affects at least 2% of the general population. It is the sixth leading cause of disability worldwide affecting the 18- to 44-year-old age group [1], a major risk factor for suicide, and the single most expensive mental health condition [2]. As BD is a chronic condition, individuals with BD often require continuous, long-term monitoring and care. However, 40% to 50% of those with any serious mental illness, including BD, do not receive any mental health treatment [3,4] despite the knowledge that evidenced-based interventions improve engagement with providers [5]. Commonly reported barriers to obtaining adequate care include lack of access, cost, and stigma [6-9]. The vast majority of patients with BD receive care in low-resource settings, which include primary care and community mental health clinics [1,2]. These providers are overburdened and lack the capacity to provide consistent monitoring shown to be effective in the management of individuals with BD [3,10-12]. Mobile health (mHealth) technologies have the potential to overcome these barriers and improve access to care [13]. There are many obstacles to implementing mobile and digital health strategies [14]. Therefore, it is imperative that the field adopts digital health strategies that are grounded in evidenced-based research; offers greater flexibility in terms of treatment options; and accommodates long-term use by heterogeneous patient populations, including varying levels of user engagement or commitment.

Existing or widely available mHealth apps for BD that are readily downloadable on smartphones typically only offer mood tracking or symptom monitoring, do not cover other important areas for individuals with BD (eg, medication adherence, sleep, and intervention), or are not supported by empirical evidence. These include apps such as eMoods Bipolar Mood Tracker, Moodlog, UP! – Depression, Bipolar & Borderline Management, and Bipolar Track. Findings of a systematic review of available apps for BD indicate that the most widely available apps do not reference clinical practice guidelines, standard psychoeducation information, or established self-management tools [15]. Other work on existing BD-related apps found that BD consumer needs were not adequately addressed by the currently available apps [16]. The top features included were mood tracking and journaling [17], and it did not provide adequate educational or other intervention content. Of the 100 top-returned publicly available apps for BD that were reviewed, only 56% mention BD in the content or description, and only 1 app was supported by peer-reviewed research. These findings suggest that mHealth tools for BD have not been widely translated into evidence-based, clinically relevant apps that can be made available to the public.

One specific mHealth tool that is empirically supported—the MONARCA app—uses sensors and self-assessments to gather information about the user's sleep, social activity, and mood, with the goal of providing information to the patients and their providers, but it does not include self-management and educational components [18]. Platforms and databases have been developed to help professionals and consumers access mHealth apps that translate evidenced-based programs and provide guidelines for consumers and help to navigate the mental health app marketplace, including some of the apps listed above [19,20]. However, very few scientifically reviewed apps specifically focus on the needs of individuals with BD, especially beyond mood tracking. One newly developed smartphone-based, self-management intervention for BD (LiveWell) was designed considering empirically supported therapy and included user input in its development, which addresses the need for empirically supported apps to provide self-management components [21]. Therefore, there is a need to focus on the development of other apps to provide options for patients with BD who would like self-management strategies and education, including support with sleep, understanding early warning signs, triggers, and maintaining healthy lifestyles [22,23].

Life Goals Collaborative Care (LGCC) [24] is one such evidence-based intervention based on the chronic collaborative care model that provides proactive care for patients through several components, including patient self-management education; care coordination across providers, predominantly through care manager contacts and improved information systems; and decision support tools for providers. LGCC's central patient-centered tenet focuses on empowering patient-self management skills through the Life Goals program, which is a series of 6 or more self-management sessions customized to individual needs and focused on mental and physical wellness, understanding symptoms, and setting personal goals. Although a component of LGCC, the Life Goals self-management program can serve as a stand-alone, manualized program for individuals with BD [25], and it has been offered via a user-friendly, provider-facing website that guides the Life Goals provider and their patient to customize an LGCC program and a consumer self-directed guidebook [26].

Self-management programs such as those included within the LGCC have been shown to improve medical and psychiatric outcomes for persons with serious mental illness after 6 months [27,28]. Several randomized trials found that LGCC reduced overall affective symptoms and improved overall role function, quality of life, participant satisfaction, and medical outcomes, compared to usual care [24,27].

Self-management programs are patient-centered and encourage regular engagement with the provider. This is challenging and burdensome for providers to implement in low-resource communities. Community-based practices lack the staffing time and capacity (eg, physical space) to offer self-management sessions such as the original LGCC program. They may also face barriers to reimbursement [29]. At the patient level, access to effective, evidence-based self-management programs are also limited outside of in-person care, which may be challenging to access owing to cost, workforce, and stigma. There is a need to develop self-management programs into mHealth technologies, with the hope that such apps will fill this gap by alleviating the burden on the providers and empower those with serious mental illness by providing them with more flexible access to evidence-based treatments.

Barriers regarding access to care and available mHealth platforms were the impetus for the development of a standalone



XSL•FO
RenderX

Life Goals self-management mobile app for persons with BD that can be accessed either in a *direct-to-consumer* format or as an augmentation to *in-person* care. The aim of this study is to evaluate the feasibility, usability, and acceptability of the Life Goals self-management app among individuals diagnosed with BD over a 6-month period. Outcome patterns of key clinical outcomes relating to functioning, disability, and psychiatric symptoms will also be assessed. The Life Goals app adds value to the existing mHealth apps for BD given its evidenced-based content, intervention effectiveness, and functionality of being customizable to the user's need.

## Methods

### Participant Recruitment

For this pre-post-test study, participants were recruited from the Heinz C. Prechter Longitudinal Study of Bipolar Disorder, an observational, naturalistic cohort study gathering phenotypic and biological data, at the University of Michigan [30], or through the UMHealthResearch web-based tool [31]. Individuals were included in the study if they were aged 18 years or older, had any BD diagnosis (ie, Bipolar I, II, or not otherwise specified [NOS], according to the Diagnostic Statistic Manual version IV [DSM-IV] criteria), were community dwelling (ie, not living in a nursing home or other institution), and were current owners and users of a smartphone. Diagnosis was confirmed using medical records or a diagnostic interview using the Diagnostic Interview for Genetics Study [32] or the Mini-International Neuropsychiatric Interview (MINI) [33]. Exclusion criteria were any serious illness precluding participation in Life Goals components as indicated by the provider, or inability to provide informed consent. One participant was excluded because she did not meet the criteria for BD after completing the MINI. Three participants discontinued from the study before completion owing to time constraints (n=2) or because they had difficulty using the app on their phone (n=1). This study was approved by the University of Michigan Internal Review Board, and all participants provided signed informed consent.

### Smartphone App

The Life Goals app was custom built to use either Apple or Android operating system. Participants were sent invitations to download the app from either Apple App Store or Google Play. Once they logged into the app, they could review the privacy policy (created by University of Michigan's Office of Technology Transfer), which was also saved under the main menu tab for easy access. Participant data were stored securely on University of Michigan's approved servers and protected by the university's security systems. A trained research associate provided training on how to use the app and provided a user manual for troubleshooting. The study team was available to answer questions via phone every business day.

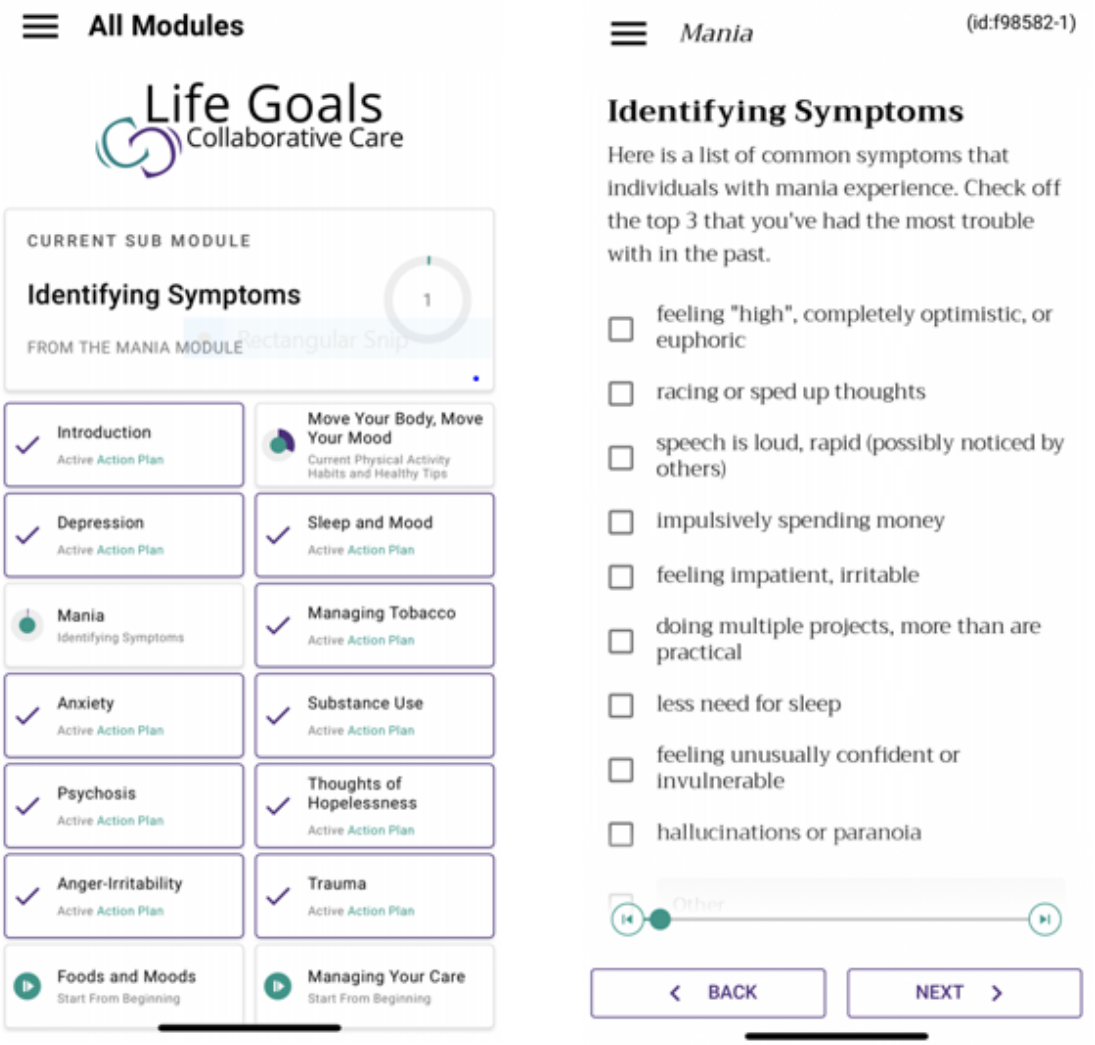
The Life Goals app includes succinct, 5- to 10-minute-long modules that provide self-management activities for managing everyday needs of individuals with BD, including mood symptom coping strategies, stigma concerns, emotional self-awareness and family support, anger and irritability, and preparation for doctor's visits. The content and wording were adapted from the original LGCC program. We encouraged participants to engage with Life Goals modules once a day, but they were ultimately self-guided, interactive modules completed at the user's own pace. Participants could type in free-text personal responses for questions asked by the app. All participants were required to complete the introduction module first, where they learned about self-management, collaborative care, stigma, value, goals, and factors impacting their mental health. Participants were also encouraged to complete the Managing Your Care module at some point, which discusses how to communicate and work with providers. If participants did not engage with the app at all within 7 days, they received one automatic notification as follows: "You haven't logged into the Life Goals app in a while. Tap here to start working on a module." If they left a module incomplete within the past 7 days and did not return to it, they received the following prompt: "You haven't completed your module. Tap here to take you back to where you left off." In all, participants received a total of 3 automated notifications for app inactivity.

Following the mandatory introductory module, the other 13 Life Goals modules covered the following mental health and wellness topics: (1) Managing Your Care; (2) Depression; (3) Mania; (4) Anxiety; (5) Trauma; (6) Thoughts of Hopelessness; (7) Psychosis; (8) Anger/Irritability; (9) Substance Use; (10) Foods and Moods; (11) Move Your Body, Move Your Mood; (12) Managing Tobacco; and (13) Sleep and Mood. Each topic is customizable based on what the participant needs or wants to learn about. Participants were encouraged to complete at least 6 modules in all. See Figure 1 for example screenshots from the Life Goals app.



XSL•FO
RenderX

**Figure 1.** Life Goals app screenshots: landing page and Identifying Triggers (Depression module).




## Feasibility Assessment

Feasibility and patient engagement were assessed through data on the number of times the participants engaged with the Life Goals app and the number of minutes they spent using the Life Goals app during the 6-month study period. A post-study Questionnaire about the user's experiences with the app was used to assess usability and acceptability and collected only at the 6-month interval. The survey contained 7 statements that participants rated their agreement with on a Likert scale, ranging from "Strongly Agree" to "Strongly Disagree." With regard to the usability of the app, participants were asked to rate their agreement with the statements "The material was displayed adequately on my phone screen" and "I had problems accessing the app…because of technical difficulties." Key statements relating to acceptability of the app as a self-management tool were, "The app improved my ability to manage my own health," and "The app helped me make progress on my wellness goals" (see Table 1). Participants were also asked a single open-ended question, in response to which they could provide us with additional free-text feedback on their app use experience.



XSL•FO
RenderX

**Table 1.** Participant ratings on poststudy evaluation (n=25).

| Evaluation question | Participants, n (%) | | |
|---|---|---|---|
| | Strongly agree or agree | Neutral | Strongly disagree or disagree |
| 1. The Life Goals app improved my ability to manage my own health. | 9 (36) | 10 (40) | 6 (24) |
| 2. The Life Goals app material was useful to me. | 15 (60) | 4 (16) | 6 (24) |
| 3. I enjoyed working through the Life Goals app material. | 9 (36) | 10 (40) | 6 (24) |
| 4. The Life Goals app helped me make progress on my wellness goals. | 10 (40) | 4 (16) | 11 (44) |
| 5. The Life Goals app was easy to use. | 18 (72) | 2 (8) | 5 (20) |
| 6. The material was displayed adequately on my phone screen. | 22 (88) | 1 (4) | 2 (8)[a] |
| 7. I had problems accessing the Life Goals app because of connectivity or other technical difficulties. | 3 (12) | 3 (12) | 19 (76) |

[a]Response: "Don't know."

### Outcome Assessment

Clinical outcomes were self-reported using validated measures of mood, health, and disability at 3 time points: baseline (study entry) and 3 and 6 months postinitiation of Life Goals app. A 6-month time span was chosen, as the original LGCC showed improvements in functioning after this time point [28]; however, in examining the feasibility of a stand-alone self-management app, we added a 3-month assessment period given that many individuals may lose interest after the first few months. A survey of the demographics (ie, age, education, race, and gender), current employment or living status, and insurance coverage was conducted only at study entry. Outcome measures assessed at all 3 waves included health-related quality of life, measured using the 12-item Short Form Health Survey (SF-12) [34]; self-reported health or disability status, measured using World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) [35]; alcohol use, measured using Alcohol Use Disorders Identification Test (AUDIT-C) [36]; drug use, measured using Drug Abuse Screening Test (DAST-10) [37]; and psychiatric symptoms, measured using the Patient Health Questionnaire–9 [38] for depression symptoms and the Internal State Scale [39] for mania or hypomania symptoms. Surveys took approximately 5-10 minutes to complete. All self-report measures were collected digitally using Research Electronic Data Capture (REDCap) tools hosted at University of Michigan [40].

### Compensation

Participants received monetary compensation for their participation in this study and to offset any data usage they may have accrued on their own mobile phones. Participants were offered US $1 for each day they used the app (defined as any interaction within the app), regardless of how many cell phone engagements they chose to use, or up to US $180 for 6 months. They also received US $10 for each survey completed, up to US $30. Maximum compensation for the 6-month study duration was US $210.

### Analyses

Univariate and bivariate statistics were used to examine all feasibility, usability, acceptability, and exploratory clinical outcomes. Feasibility was assessed by examining usage (defined as the number of instances the app was used and the duration while using the app and its modules) across the total 6-month study duration, from baseline to 3 months, and from 3 months to 6 months. For usability and acceptability estimates, frequencies of responses and open-ended qualitative responses were reviewed and summarized to determine common themes. For our exploratory clinical outcomes, paired *t*-tests were performed to evaluate over-time change. Spearman rho correlations were performed to determine associations between over-time change in outcome measures and app usage.

## Results

### User Demographics

Of the 496 email invitations sent, 28 (5.6%) participants consented and enrolled in this study. The majority of participants were diagnosed with bipolar type I (n=21, 75%), with a smaller proportion diagnosed with bipolar type II (n=4, 14.3%) and bipolar, NOS (n=3, 10.7%). The average age of participants was 44.7 (range 24-72) years; 68% (19/28) identified as female; 75% (21/28) reported having a college degree; and 10.7% (3/28) identified as racial minorities (Multimedia Appendix 1).

### User Engagement

In this pilot study, app use waxed and waned over the course of the study. Of the 28 enrolled participants, 26 (93%) used the app at least once during the study period; 24 (86%) used the app during the first 3 months, and this number decreased to 18 (64%) during the next 3 months.

Over the full study duration, participants engaged with the app for a median of 25 times (IQR 13-65.75) and used the app for an average of 154 (IQR 39-194) minutes. However, this usage was not evenly distributed over time. During the first 3 months, participants engaged with the app for a median of 16.5 (IQR 12.25-41.25) times and for a median of 117 (IQR 22-132) minutes. During the next second 3 months, the median app use had increased to 23 (IQR 6-64.50) times, but the average time spent decreased to 66 minutes (IQR 22, 79) during the second period.

Among the subset of 18 participants who used the app at least once during the first and last 3-month periods, they also showed more app use within the first 3 months (median 11, IQR 7-28



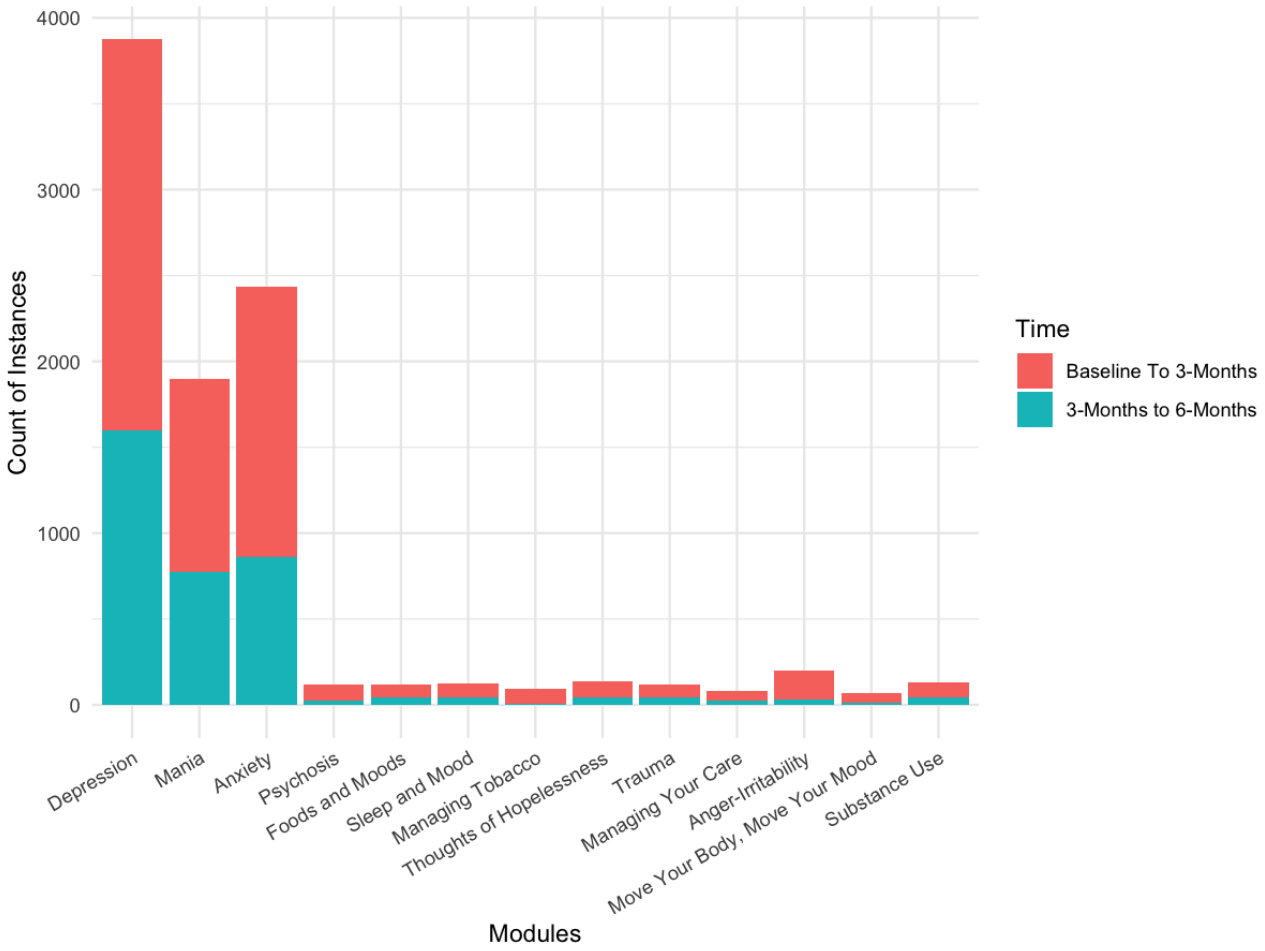
XSL•FO
RenderX

instances) than in the last 3 months (median 8, IQR 5-35 instances), which was a statistically significant reduction (P=.005). This subset also used the app for longer, on average, in the first 3 months of the study (median 7, IQR 1-59 minutes per usage) compared to the last 3 months (median 3.4, IQR 1-31 minutes; P=.006).

## Module Completion

Participants were required to complete the introduction module before attempting any other module. Furthermore, they were encouraged to use any or all modules, as deemed interesting, during the 6-month period. Of the 13 modules, excluding the mandatory introduction module, 3 (17%) participants used all 13 modules in the app. Over the course of the study, participants completed a number of modules (mean 8.2, SD 3.3; median 7; mode 5; range 4-13). All participants completed more than 3 Life Goals app modules, and 67% (12/18) of the participants completed 6 or more modules The most frequently assessed modules were the same in both study periods (ie, first 3 months and next 3 months); these included Depression, Anxiety, and Mania (Figure 2).

**Figure 2.** Number of times the Life Goals app was used at different time points.




## Usability and Acceptability of Life Goals App

A total of 25 (89%) participants completed the poststudy survey about acceptability and usability (Table 1). Of these 25 participants, the majority found the material provided by the Life Goals app useful (n=15, 60%); they found the app easy to use (n=18, 72%); and they found the screen display adequate (n=22, 88%). Few participants reported problems accessing the Life Goals app because of connectivity or other technical difficulties (3/25, 12%). However, positive impacts of the app on health management were less widely endorsed: 10 (40%) users reported finding the app helpful in managing their health, and 9 (36%) reported finding it helpful in making progress on their wellness goals.

The open-ended question where participants could provide additional feedback on their app use experience included the following responses:

> *A lot of the information was obvious to me.*
>
> *The app is well designed. I can see it being highly beneficial to someone learning how to manage their illness.*
>
> *I'm pretty stable and didn't really find the app helpful…If I was experiencing symptoms, maybe more so.*
>
> *I think it will help beginner sufferers more. My habits didn't change a lot, but Life Goals App really got me thinking differently.*



XSL•FO
RenderX

### Outcome Assessment

Paired sample *t*-tests did not show significant changes in symptoms or functioning when comparing baseline assessments to 3-month follow-up assessments, or when comparing baseline assessments to 6-month follow-up assessments (Multimedia Appendix 2), likely due to the small sample size. Effect sizes were all small to negligible (Cohen *d* range –0.372 to 0.183). However, trends for improvements in physical health functioning, mental health functioning, and mania-related well-being were noted. There were no significant relationships between time spent using the app (minutes) or the number of instances using the app within the first 3 months or the last 3 months and changes in outcome measurements during the same periods, except for a positive correlation between change in mania symptoms and Life Goals app duration (see Multimedia Appendix 3).

## Discussion

### Principal Results

This pilot study examined the feasibility, usability, and acceptability of the Life Goals self-management app among individuals diagnosed with BD. We found that the median number of times that individuals used this app was 25, for a total of 154 minutes across 6 months. As with usage patterns of other apps, the app was used for a longer time during the first 3 months than the last 3 months of the study period. In a clinical setting, Life Goals participants are expected to complete 6 sessions, or 6 Life Goals modules; completion of 3 sessions is regarded as a clinically significant dose [29,41,42]. As such, all participants completed at least 3 Life Goal app modules, and the majority completed at least 6 Life Goals modules, indicating all received a clinically significant dose comparable to Life Goals in a clinical setting. Furthermore, the number of modules completed (mean 8.2, median 7) indicated reasonable use of the modules. Not counting the introduction module, which was mandatory, the most frequently accessed modules in our sample were Depression, followed by Anxiety and Mania, suggesting that these are areas that appealed to the sample the most or directly addressed self-management needs.

For usability and acceptability, frequencies of responses from participants indicated high usability and satisfaction with the user interface. The majority of participants experienced no difficulties using the app and felt that the material was displayed adequately; however, the results showed low success in the app encouraging self-management of their own health. Only a minority of the participants felt that the app helped them to make progress on their wellness goals and improved their ability to manage their own health. Similarly, when symptoms, health, and functioning were rated over the course of the study using self-report surveys, we found no significant improvements in ratings. However, there were notable trends for mania and mental and physical health functioning over the first 3 months, using this small sample that were pointing in the direction of improved scores. Written responses from participants centered around themes related to liking the app, but the content or self-management pieces were not novel to them, and a few suggested that they felt that the app would be more beneficial to individuals who were newly diagnosed. These findings suggest that targeting the app to those who have received less psychoeducation about BD or were earlier in their illness course may find greater benefit. Future work to understand what individuals with BD look for in similar apps is needed. Only 6% of all those who offered to participate in this study agreed to use the app, suggesting that there may be specific factors that influence the uptake of this type of app. Our future work aims to explore both dissemination via direct-to-consumer and clinically integrated pathways that would address the question about uptake and engagement with self-management apps, such as Life Goals. Recent work suggesting design considerations for development, engagement, and evaluation [43] in apps for BD will be helpful to refine the development and dissemination of this work.

### Limitations

Given our small sample size, our group of participants with BD may not be reflective of the broader, heterogeneous nature of BD. Furthermore, it is possible that a larger sample would have sufficient power to detect changes in clinical outcomes. We did not include a control group, so it was not possible to determine whether clinical changes were related to use of the Life Goals app. Our study design likely resulted in the tendency to include individuals who are more technology oriented and motivated to participate in research, as well as those who are more willing to have an almost daily engagement with the Life Goals app. For example, our sample was already actively engaged in a longitudinal study of BD, so these participants are likely more willing to engage in other research studies and participate in daily activities than are other individuals who are not volunteering their time. Future studies will need to evaluate the use of the Life Goals app in a more rigorous or controlled manner and within a more generalizable population to examine whether use statistics were related to incentivizing participants or whether symptom improvement trends persist over time. Lastly, we compensated our participant's time in this study, which may have increased their interest to participate or their motivation to engage with the Life Goals app.

### Conclusions

Overall, these results indicate that the Life Goals app, a smartphone self-management app developed using an evidence-based intervention based on the chronic collaborative care model, is feasible and acceptable for individuals with BD. Individuals using the Life Goals app may use it more frequently at first and may have the greatest interest in mood-related modules rather than other wellness modules. Wider dissemination of the app to individuals in different stages of recovery or earlier in the illness is needed. This app shows potential as an mHealth technology based on evidence-based treatments that can help bridge the gap in access to care and reducing burden on providers. It can also offer greater flexibility in terms of treatment options that accommodates varying levels of user engagement and commitment.



XSL•FO
RenderX

## Acknowledgments

This project was funded by the Michigan Department of Health and Human Services, The Heinz C Prechter Research Program and the Richard Tam Foundation.

## Conflicts of Interest

MGM has consulted for Otsuka and Janssen Pharmaceuticals and receives research support from Janssen Pharmaceuticals. MGM has also consulted with the Milken Institute for Strategic Philanthropy. AMK and CL are named on the Life Goals app invention disclosure filed with the University of Michigan. AMK is a co-author of a book about Life Goals and receives royalties from New Harbinger publications. SNS, KAR, AKY, HB, IC, BN, and EV has no conflict of interest to declare.

Multimedia Appendix 1
Demographic characteristics of the study participants (N=28).
[DOCX File , 19 KB - formative_v5i12e32450_app1.docx ]

Multimedia Appendix 2
Self-report clinical outcome data with completion rate, mean score, and SD at each time point.
[DOCX File , 15 KB - formative_v5i12e32450_app2.docx ]

Multimedia Appendix 3
Correlations between Life Goals app use and outcome measures for time period 1 (ie, baseline to 3-month follow-up) and time period 2 (ie, 3-month to 6-month follow-up).
[DOCX File , 94 KB - formative_v5i12e32450_app3.docx ]

## References

1. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: global burden of disease study. Lancet 1997 May 17;349(9063):1436-1442. [doi: 10.1016/S0140-6736(96)07495-8] [Medline: 9164317]
2. Peele PB, Xu Y, Kupfer DJ. Insurance expenditures on bipolar disorder: clinical and parity implications. Am J Psychiatry 2003 Jul;160(7):1286-1290. [doi: 10.1176/appi.ajp.160.7.1286] [Medline: 12832243]
3. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol. Med 2010 Dec 07;41(8):1751-1761. [doi: 10.1017/s0033291710002291]
4. Kreyenbuhl J, Nossel IR, Dixon LB. Disengagement from mental health treatment among individuals with schizophrenia and strategies for facilitating connections to care: a review of the literature. Schizophr Bull 2009 Jul 02;35(4):696-703 [FREE Full text] [doi: 10.1093/schbul/sbp046] [Medline: 19491314]
5. Dixon LB, Holoshitz Y, Nossel I. Treatment engagement of individuals experiencing mental illness: review and update. World Psychiatry 2016 Feb 01;15(1):13-20. [doi: 10.1002/wps.20306]
6. Huxley CJ, Atherton H, Watkins JA, Griffiths F. Digital communication between clinician and patient and the impact on marginalised groups: a realist review in general practice. Br J Gen Pract 2015 Dec;65(641):e813-e821 [FREE Full text] [doi: 10.3399/bjgp15X687853] [Medline: 26622034]
7. Pepin R, Segal DL, Coolidge FL. Intrinsic and extrinsic barriers to mental health care among community-dwelling younger and older adults. Aging Ment Health 2009 Sep;13(5):769-777. [doi: 10.1080/13607860902918231] [Medline: 19882416]
8. Rowan K, McAlpine DD, Blewett LA. Access and cost barriers to mental health care, by insurance status, 1999-2010. Health Aff (Millwood) 2013 Oct;32(10):1723-1730 [FREE Full text] [doi: 10.1377/hlthaff.2013.0133] [Medline: 24101061]
9. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. Health Aff (Millwood) 2009 Jun;28(3):w490-w501 [FREE Full text] [doi: 10.1377/hlthaff.28.3.w490] [Medline: 19366722]
10. Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RMA, Petukhova M, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Arch Gen Psychiatry 2007 May 01;64(5):543-552 [FREE Full text] [doi: 10.1001/archpsyc.64.5.543] [Medline: 17485606]
11. Cerimele JM, Fortney JC, Unützer J. Bipolar disorder and population health. Psychiatr Serv 2017 Feb 01;68(2):192-194 [FREE Full text] [doi: 10.1176/appi.ps.201600011] [Medline: 27524370]
12. Macdonald A, Adamis D, Craig T, Murray R. Continuity of care and clinical outcomes in the community for people with severe mental illness. Br J Psychiatry 2019 May 05;214(5):273-278. [doi: 10.1192/bjp.2018.261] [Medline: 31012407]
13. Sood M, Chadda RK, Singh P. Mobile health (mHealth) in mental health: scope and applications in low-resource settings. Natl Med J India 2016;29(6):341-343 [FREE Full text] [Medline: 28327483]
14. Morton E, Torous J, Murray G, Michalak EE. Using apps for bipolar disorder – An online survey of healthcare provider perspectives and practices. Journal of Psychiatric Research 2021 May;137:22-28. [doi: 10.1016/j.jpsychires.2021.02.047]
15. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]



XSL•FO
RenderX

16. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
17. Lagan S, Ramakrishnan A, Lamont E, Ramakrishnan A, Frye M, Torous J. Digital health developments and drawbacks: a review and analysis of top-returned apps for bipolar disorder. Int J Bipolar Disord 2020 Dec 01;8(1):1-8. [doi: 10.1186/s40345-020-00202-4]
18. Matthews M, Abdullah S, Murnane E, Voida S, Choudhury T, Gay G, et al. Development and evaluation of a smartphone-based measure of social rhythms for bipolar disorder. Assessment 2016 Aug 29;23(4):472-483 [FREE Full text] [doi: 10.1177/1073191116656794] [Medline: 27358214]
19. One Mind PsyberGuide – A mental health app guide. 2021 Nov 24. URL: https://onemindpsyberguide.org/ [accessed 2021-11-24]
20. M-health index & navigation database 2020. URL: https://mindapps.org/ [accessed 2021-11-15]
21. Jonathan GK, Dopke CA, Michaels T, Bank A, Martin CR, Adhikari K, et al. A smartphone-based self-management intervention for bipolar disorder (livewell): user-centered development approach. JMIR Ment Health 2021 Apr 12;8(4):e20424 [FREE Full text] [doi: 10.2196/20424] [Medline: 33843607]
22. Nicholas J, Boydell K, Christensen H. Beyond symptom monitoring: consumer needs for bipolar disorder self-management using smartphones. Eur Psychiatry 2017 Jul 23;44:210-216. [doi: 10.1016/j.eurpsy.2017.05.023] [Medline: 28692910]
23. Switsers L, Dauwe A, Vanhoudt A, Van Dyck H, Lombaerts K, Oldenburg J. Users' perspectives on mhealth self-management of bipolar disorder: qualitative focus group study. JMIR Mhealth Uhealth 2018 May 02;6(5):e108 [FREE Full text] [doi: 10.2196/mhealth.9529] [Medline: 29720363]
24. Bauer MS, McBride L, Williford WO, Glick H, Kinosian B, Altshuler L, Cooperative Studies Program 430 Study Team. Collaborative care for bipolar disorder: part I. Intervention and implementation in a randomized effectiveness trial. Psychiatr Serv 2006 Jul;57(7):927-936. [doi: 10.1176/ps.2006.57.7.927] [Medline: 16816276]
25. Kilbourne AM, Post EP, Nossek A, Sonel E, Drill LJ, Cooley S, et al. Service delivery in older patients with bipolar disorder: a review and development of a medical care model. Bipolar Disord 2008 Sep;10(6):672-683. [doi: 10.1111/j.1399-5618.2008.00602.x] [Medline: 18837861]
26. Bauer M, Kilbourne A, Greenwald D, Ludman E. Overcoming bipolar disorder: a comprehensive workbook for managing your symptoms and achieving your life goals. Overcoming bipolar disorder: a comprehensive workbook for managing your symptoms and achieving your life goals: New Harbinger Publications; 2009:a.
27. Kilbourne AM, Post EP, Nossek A, Drill L, Cooley S, Bauer MS. Improving medical and psychiatric outcomes among individuals with bipolar disorder: a randomized controlled trial. Psychiatr Serv 2008 Jul;59(7):760-768. [doi: 10.1176/ps.2008.59.7.760] [Medline: 18586993]
28. Kilbourne AM, Goodrich DE, Lai Z, Clogston J, Waxmonsky J, Bauer MS. Life Goals Collaborative Care for Patients With Bipolar Disorder and Cardiovascular Disease Risk. PS 2012 Dec;63(12):1234-1238. [doi: 10.1176/appi.ps.201100528]
29. Kilbourne AM, Almirall D, Eisenberg D, Waxmonsky J, Goodrich DE, Fortney JC, et al. Protocol: Adaptive Implementation of Effective Programs Trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. Implement Sci 2014 Sep 30;9(1):132-114 [FREE Full text] [doi: 10.1186/s13012-014-0132-x] [Medline: 25267385]
30. McInnis M, Assari S, Kamali M, Ryan K, Langenecker S, Saunders E. Cohort Profile: the Heinz C. Prechter longitudinal study of bipolar disorder. International journal of epidemiology. n 2018;47(1):28. [doi: 10.1093/ije/dyx229]
31. UMHealthResearch. URL: https://umhealthresearch.org/ [accessed 2021-11-24]
32. Nurnberger JJ, Blehar M, Kaufmann C, York-Cooler C, Simpson S, Harkavy-Friedman J. Diagnostic Interview for Genetic Studies (DIGS). APA PsycTests 1994;51(11):63-59. [doi: 10.1037/t64336-000]
33. Sheehan D, Lecrubier Y, Sheehan K, Amorim P, Janavs J, Weiller E. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998:22-33. [doi: 10.1037/t18597-000]
34. Ware J, Kosinski M, Keller S. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003]
35. Chisolm TH, Abrams HB, McArdle R, Wilson RH, Doyle PJ. The WHO-DAS II: psychometric properties in the measurement of functional health status in adults with acquired hearing loss. Trends Amplif 2005 Aug 23;9(3):111-126 [FREE Full text] [doi: 10.1177/108471380500900303] [Medline: 16244758]
36. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
37. Lam LP, Leung WC, Ip P, Chow CB, Chan MF, Ng JWY, et al. Validation of the Drug Abuse Screening Test (DAST-10): a study on illicit drug use among Chinese pregnant women. Sci Rep 2015 Jun 19;5(1):11420-11411 [FREE Full text] [doi: 10.1038/srep11420] [Medline: 26091290]
38. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]



XSL•FO
RenderX

39. Bauer MS, Vojta C, Kinosian B, Altshuler L, Glick H. The Internal State Scale: replication of its discriminating abilities in a multisite, public sector sample. Bipolar Disord 2000 Dec 24;2(4):340-346. [doi: 10.1034/j.1399-5618.2000.020409.x] [Medline: 11252648]
40. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
41. Smith SN, Almirall D, Prenovost K, Liebrecht C, Kyle J, Eisenberg D, et al. Change in patient outcomes after augmenting a low-level implementation strategy in community practices that are slow to adopt a collaborative chronic care model: a cluster randomized implementation trial --II. Med Care 2019 Jul;57(7):503-511 [FREE Full text] [doi: 10.1097/MLR.0000000000001138] [Medline: 31135692]
42. Smith SN, Liebrecht CM, Bauer MS, Kilbourne AM. Comparative effectiveness of external vs blended facilitation on collaborative care model implementation in slow-implementer community practices. Health Serv Res 2020 Dec 30;55(6):954-965. [doi: 10.1111/1475-6773.13583] [Medline: 33125166]
43. Morton E, Nicholas J, Lapadat L, O'Brien HL, Barnes SJ, Poh C, et al. Use of smartphone apps in bipolar disorder: an international web-based survey of feature preferences and privacy concerns. J Affect Disord 2021 Dec 01;295:1102-1109. [doi: 10.1016/j.jad.2021.08.132] [Medline: 34706421]

## Abbreviations

**Audit-C:** Alcohol Use Disorders Identification Test
**BD:** bipolar disorder
**DAST-10:** Drug Abuse Screening Test
**DSM-IV:** Diagnostic Statistic Manual version IV
**LGCC:** Life Goals Collaborative Care
**mHealth:** mobile health
**NOS:** not otherwise specified
**REDCap:** Research Electronic Data Capture
**SF-12:** 12-item Short Form Health Survey
**WHO-DAS:** World Health Organization Disability Assessment Schedule 2.0

*Edited by G Eysenbach; submitted 30.07.21; peer-reviewed by I Choi; comments to author 26.08.21; revised version received 08.09.21; accepted 05.11.21; published 13.12.21.*

*Please cite as:*
*Ryan KA, Smith SN, Yocum AK, Carley I, Liebrecht C, Navis B, Vest E, Bertram H, McInnis MG, Kilbourne AM*
*The Life Goals Self-Management Mobile App for Bipolar Disorder: Consumer Feasibility, Usability, and Acceptability Study*
*JMIR Form Res 2021;5(12):e32450*
*URL: https://formative.jmir.org/2021/12/e32450*
*doi:10.2196/32450*
*PMID:34898452*

©Kelly A Ryan, Shawna N Smith, Anastasia K Yocum, Isabel Carley, Celeste Liebrecht, Bethany Navis, Erica Vest, Holli Bertram, Melvin G McInnis, Amy M Kilbourne. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# User Reviews of Depression App Features: Sentiment Analysis

Julien Meyer[1], MSc, PhD; Senanu Okuboyejo[2], PhD

[1]School of Health Services Management, Ted Rogers School of Management, Ryerson University, Toronto, ON, Canada
[2]Department of Computer and Information Science, Covenant University, Ota, Nigeria

**Corresponding Author:**
Julien Meyer, MSc, PhD
School of Health Services Management
Ted Rogers School of Management
Ryerson University
350 Victoria Street
Office: TRS 3-081
Toronto, ON, M4T 1G6
Canada
Phone: 1 4169795000 ext 6296
Email: julien.meyer@ryerson.ca

## Abstract

**Background:** Mental health in general, and depression in particular, remain undertreated conditions. Mobile health (mHealth) apps offer tremendous potential to overcome the barriers to accessing mental health care and millions of depression apps have been installed and used. However, little is known about the effect of these apps on a potentially vulnerable user population and the emotional reactions that they generate, even though emotions are a key component of mental health. App reviews, spontaneously posted by the users on app stores, offer up-to-date insights into the experiences and emotions of this population and are increasingly decisive in influencing mHealth app adoption.

**Objective:** This study aims to investigate the emotional reactions of depression app users to different app features by systematically analyzing the sentiments expressed in app reviews.

**Methods:** We extracted 3261 user reviews of depression apps. The 61 corresponding apps were categorized by the features they offered (*psychoeducation*, *medical assessment*, *therapeutic treatment*, *supportive resources*, and *entertainment*). We then produced word clouds by features and analyzed the reviews using the Linguistic Inquiry Word Count 2015 (Pennebaker Conglomerates, Inc), a lexicon-based natural language analytical tool that analyzes the lexicons used and the valence of a text in 4 dimensions (*authenticity*, *clout*, *analytic*, and *tone*). We compared the language patterns associated with the different features of the underlying apps.

**Results:** The analysis highlighted significant differences in the sentiments expressed for the different features offered. Psychoeducation apps exhibited more *clout* but less authenticity (ie, personal disclosure). *Medical assessment* apps stood out for the strong negative emotions and the relatively negative ratings that they generated. *Therapeutic treatment* app features generated more positive emotions, even though user feedback tended to be less authentic but more analytical (ie, more factual). *Supportive resources* (connecting users to physical services and people) and *entertainment* apps also generated fewer negative emotions and less anxiety.

**Conclusions:** Developers should be careful in selecting the features they offer in their depression apps. *Medical assessment* features may be riskier as users receive potentially disturbing feedback on their condition and may react with strong negative emotions. In contrast, offering information, contacts, or even games may be safer starting points to engage people with depression at a distance. We highlight the necessity to differentiate how mHealth apps are assessed and vetted based on the features they offer. Methodologically, this study points to novel ways to investigate the impact of mHealth apps and app features on people with mental health issues. mHealth apps exist in a rapidly changing ecosystem that is driven by user satisfaction and adoption decisions. As such, user perceptions are essential and must be monitored to ensure adoption and avoid harm to a fragile population that may not benefit from traditional health care resources.

***(JMIR Form Res 2021;5(12):e17062)*** doi:10.2196/17062



XSL•FO
RenderX

**KEYWORDS**

mHealth; depression; app reviews; natural language processing; app features; emotions; use; linguistic inquiry word count; mobile phone

## Introduction

### Background

Major depressive disorders account for almost 300 million cases worldwide, with a loss of 63 million disability-adjusted life years every year [1]. Effective treatments are frequently unavailable to those with the greatest need [2,3]. Barriers to receiving mental health and behavioral care include transportation problems, time constraints, costs, emotional barriers, and stigma [3]. Young adults, in particular, tend to have a negative opinion of the mental health care system, feel disconnected from its services, and prefer handling their concerns by themselves rather than resorting to mental health care services [4].

Mobile health (mHealth) apps deliver health care through mobile information technologies such as smartphones and offer an opportunity to address these barriers and expand the reach of depression care, especially in areas with limited or no specialists. They offer multiple advantages such as quasi-unlimited capacity, 24/7 availability, equitable access, anonymity, tailored approach, links to other systems, and low cost [5]. Patients also tend to prefer psychological treatment to medication [4]. As a result, mHealth apps have been installed millions of times [6]. mHealth apps can address a variety of needs and researchers have identified 6 categories of features offered by depression apps, as summarized in Textbox 1 [6,7].

**Textbox 1.** Definition of depression apps features.

Psycho-education

- Educate, train or inform users through books, guides, news, journal articles, commentaries or opinions, tips, and lessons.

Medical assessment

- Allow users to screen, diagnose, assess risks, assess self, determine treatment.

Symptom management

- Allow users to track symptoms, gather history, including physical health data and provide useful, comprehensible output.

Therapeutic treatment

- Prescribe solutions to improve the condition (therapeutic or not). Includes relaxation, hypnosis, mindfulness exercises, meditation, spiritual or faith-based solutions.

Supportive resources

- Provide referrals for help, connects users with support, for example, emotional and social support, treatment interventions for acute or chronic use, etc.

Entertainment

- Serve recreational purposes, such as quotes, dark humor, wall papers and games.

Multifeature

- Offer 2 or more of the above features.

Evidence suggests that mHealth apps can be effective for various mental health disorders, including depression [8-12]. Nevertheless, evidence remains scarce and incomplete. First, there are a limited number of studies and participants involved [9] and they rarely focus on the benefits and usefulness of specific features [6,7]. Second, the rapid evolution of these apps implies that the investment in clinical evaluations of specific apps may be short-lived; for instance, a study showed that 50% of the depression apps available were different after 130 days, with a new app for depression being available every 3 days [13].

Moreover, the clinical relevance seems to be unrelated to adoption [13]. App developers indicate that app stores are increasingly their favorite distribution channel, over health care providers, in line with the evolution of user decision patterns toward increased patient empowerment [14]. Young users, in particular, often openly reject being told what app to use [15]. Thus, adoption of mHealth apps is increasingly driven by patient attitudes about and experiences with the apps than by clinical evidence and professionals' prescriptions. Therefore, to better address user needs and improve the adoption and use of adequate apps, we need to better understand the experiences, behaviors, and attitudes of the users of these mHealth apps [16].

App reviews are one key source of information on apps. App users are invited by app stores to rate the apps they have installed and write free text reviews about them. Users willingly and spontaneously contribute to these reviews. Although only



XSL•FO
RenderX

a fraction of users choose to do so, these reviews aggregate into a large data set of mostly authentic and publicly available data on user experience with depression apps.

These app reviews are important because of several reasons. First, app reviews can help identify bugs, user requirements, feature requests, user experience with specific app features [17,18] and whether the existing features meet user expectations [19]. Consumer knowledge also reflects different types of knowledge that may be expressed in the reviews, such as knowledge of attributes, the topic, or the buying process [20].

Second, user reviews inform us about the experiences and mental states of the users. For instance, the choice of words in web-based blogs by people with mental health issues helped identify young adults' suspicions toward the mental health system [4].

Finally, user reviews influence the prospective users' decision to choose a health care service or not [21]. Specifically, the number of user reviews and user ratings of mental health apps influence adoption [22] and the expression of emotions in user reviews, notably negative emotions, influences how people interpret the reviews [23].

As the scope of human endeavors supported by technology continues to broaden and become more intimate, emotions and values tend to play an increasing role in explaining the adoption and use of a technology [24,25]. Users' emotional reactions can reflect their assessment of an app, their future propensity to use them, and also reflect some of the impacts of the apps on the users' condition and behavior.

### Objectives

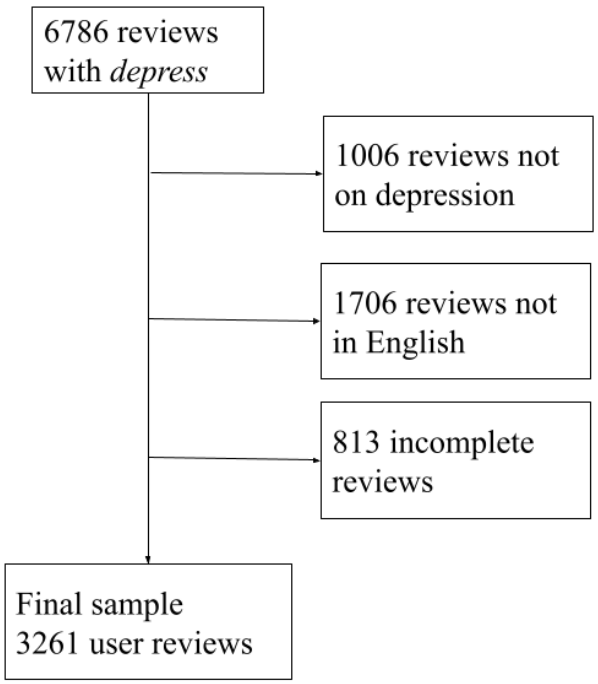
Considering the importance of user emotions and user reviews in understanding user decisions, this study investigates the emotional reactions expressed in user reviews of depression health apps and analyzes how they relate to app features.

## Methods

### Data Collection

Depression apps and their reviews were scraped from Google Play Store and Apple App Store through 42Matters, a third-party application programming interface provider, based on all apps worldwide that included the root *depress*- either in the title or in the description. The search was conducted in March 2018. The data set was then cleaned manually by a researcher to remove non–English-language app reviews or apps and apps unrelated to the mental condition (eg, depression used as a geological term) and apps with missing data (Figure 1). Screening of the app reviews resulted in a final data set containing 3261 app reviews associated with 61 apps.

**Figure 1.** Diagram of app reviews selection process.




Each app was categorized using functional categories defined in other studies on depression apps [6,7]. For each app, a researcher read the description of the app, installed the app if necessary, and coded the feature or features provided by the app (refer to Textbox 1 for the coding scheme).

### Statistical Analysis

In app reviews, users report and document their experiences in an unstructured and nonmethodical manner. The volume, variety, velocity, and veracity of user reviews also contributed to making them difficult to analyze.

As a first step, we created a word cloud of the content of the app reviews by functional category. This provided an intuitive, unmediated idea of some of the themes and concepts reflected in the reviews. We used the statistical and data management software R to remove common English stop words (plus *depression*,*depressed* and *app*) and used the 50 most frequently used words for each group of app reviews (refer to Multimedia



XSL•FO
RenderX

Appendix 1 for the list and frequencies of the most frequently used words by feature).

We then used a text-mining tool to analyze the content of the reviews. Natural language processing involves techniques to analyze large data sets of natural (ie, not codified) language and has been critical for understanding consumer attributes and behaviors [26,27]. Sentiment analysis is a subset of natural language processing that investigates thoughts, emotional reactions, and feelings regarding a specific subject or topic or simply identifies the overall polarity of a topic [21]. It operates by extracting and retrieving information from unstructured raw text and extracting words or grammatical patterns that reflect emotions or thought processes. In health care, these analytical methods have been used, for instance, to interpret textual information about patient experience [21,28] or patient satisfaction [29].

App reviews were analyzed using Linguistic Inquiry Word Count (LIWC) 2015 (Pennebaker Conglomerates Inc). LIWC is a well-established application that analyzes natural language text segments and counts the frequency of words reflecting different emotions, thinking styles, social concerns, and other dimensions [30-32]. LIWC is a lexicon-based approach to semantic analysis, which is based on a predefined dictionary. Although LIWC was not specifically developed to investigate app reviews, lexicon-based approaches to sentiment analysis of consumer reviews do not significantly differ based on the context being analyzed [33]. LIWC has been used in prior studies to extract depression-related linguistic cues from web-based forums [31] and analyze mobile app reviews [19,34].

LIWC codifies more than 92 different aspects of language. It assesses the valence of a text in 4 dimensions (*authenticity*, *clout*, *analytic*, *tone*) by analyzing the linguistic style. Authenticity measures the presence of features associated with true and false stories [32,35]. False stories, for instance, tend to use more motion words and more negative emotion words but fewer first-person pronouns [35]. LIWC then provides a rating from 0 to 100, 50 being neutral. For example, the following review was scored 99 on authenticity, 73.64 on (emotional) tone, but only 1 on *analytic* and *clout*, reflecting that the quote talked more about the user's experience than about the app:

> *I already knew this but now I can physically see that I am and I can't even tell my own parents wow. I don't know how to get better I really don't and it said I have server depression.*

In contrast, the following review was rated 64.27 on *analytic*, 98.93 on *clout*, reflecting how the user was analyzing the features and trying to influence the designers but not saying much about his or her experience with the app:

> *Would Love to have Transparent Effect! And an idea for you, make one with Motivational thoughts and people would go CRAZY and Install your App.*

LIWC also measures the frequency of certain lexicons, such as *money*, *home*, *you* or *adverb*, (not only these exact words but any words related to the theme). Categories were rated from 0 to 100 to reflect the valence in the linguistic feature, 0 indicating complete absence and 100 indicating that the fragment fully reflected the category. Owing to the purpose of the apps, the codes may reflect the user's state of mind or the feature being assessed. "Best anxiety tool out there" and "It's OK but too confusing" were both rated at 20 on anxiety but the first one reflected an analytical stance on anxiety while the other reflected the state of mind of the user. Thus, both meanings were included in the values and could not be disentangled by LIWC.

We coded the app reviews with LIWC, meaning that the complete app reviews were analyzed and rated rather than the individual sentences. Using R, we then performed 2-tailed *t* tests on the relevant dimensions to measure whether there were significant differences between the reviews associated with 1 feature and the depression app reviews overall. We focused on the 4 summary language variables (*analytic*, *clout*, *authenticity*, and *emotional tone*) [32]. Owing to its importance, instead of reporting the emotional tone directly, we reported its subcomponents *positive emotion* and *negative emotion* which have been associated with app adoption [3]. We also added *anxiety*, a subcomponent of *negative emotion*, which was directly related to depression. *Negative emotion*, *positive emotion*, and *anxiety* are lexicon dimensions that reflect frequency rather than valence. Therefore, their values were typically lower than those of the dimensions. Textbox 2 defines the variables that were retained.



XSL•FO
RenderX

**Textbox 2.** Selected Linguistic Inquiry Word Count dimensions and their definitions.

Analytic (analytical thinking)

- Degree to which people use words that suggest formal, logical, and hierarchical thinking patterns. People low in analytical thinking tend to write and think using language that is more narrative, focusing on the here-and-now and personal experiences.

Clout

- Relative social status, confidence, or leadership skill that people display through their writing or talking.

Authenticity

- When people reveal themselves in an authentic or honest manner, they are more personal, humble, and vulnerable.

Tone (emotional tone)

- Includes both positive and negative emotion dimensions; the higher the number, the more positive the tone. Ratings below 50 suggest a more negative emotional tone. It was broken down into:

Positive emotion

- The more that people use positive emotion words, the more optimistic they tend to be. If you feel good about yourself, you are more likely to see the world in a positive way.

Negative emotion

- Use of negative emotion words is weakly linked to people's ratings of anxiety or even neurotic. People who have had a bad day are more likely to see the world through negatively-tinted glasses. Words denoting anxiety (worried, fearful...) are a subset of negative emotion.

Finally, we illustrated the analysis with samples of complete reviews to connect both the word clouds and the language variables with actual uses of the words, as the anecdotal context facilitated the understanding of the analytical process.

# Results

## App Statistics

First, we present basic descriptive statistics on the number of reviews by year of publication (Table 1) and by word count (Table 2).

**Table 1.** Number of reviews by year of publication.

| Year of review | Reviews, n |
|---|---|
| 2012 | 4 |
| 2013 | 5 |
| 2014 | 18 |
| 2015 | 122 |
| 2016 | 138 |
| 2017 | 1760 |
| 2018 | 877 |
| 2019 | 337 |

**Table 2.** Number of reviews by word count.

| Word count | Reviews, n |
|---|---|
| <5 | 645 |
| 6 to 10 | 1251 |
| 11 to 20 | 807 |
| 21 to 50 | 868 |
| 50 to 100 | 292 |
| >100 | 43 |



XSL•FO
RenderX

Second, we analyzed the number of app reviews by category and the average review rating out of 5 stars (Table 3). No app that offered the feature of *symptom management* exclusively was identified, and therefore, that category was excluded from the remainder of the analysis. The app ratings for *psychoeducation*, *therapeutic treatment*, and *multifeature* apps were slightly but significantly above average, while those for *medical assessment* apps were significantly below average. The average character count was 134, which was slightly above the average of 117 observed in app reviews in general [36].

**Table 3.** App count, installation, and reviews by functional category.

| App category | Exclusive feature app count (total apps with feature; N=61), n (%) | App review count (N=3261), n (%) | Rating, mean (SD) | *P* value[a] | Review length in words, mean (SD) |
|---|---|---|---|---|---|
| *Psychoeducation* [b] | 17 (27.86) | 259 (7.94) | 4.2 (1.3) | .06 | 18.9 (18.2) |
| *Medical assessment* | 12 (19.67) | 556 (17.05) | 4.0 (1.4) | <.001 | 14.8 (15.7) |
| *Symptom management* | 0 (0) | 0 (0) | N/A[c] | N/A | N/A |
| *Therapeutic treatment* | 9 (14.75) | 293 (8.98) | 4.3 (1.3) | .69 | 18.1 (19.7) |
| *Supportive resources* | 4 (6.56) | 138 (4.23) | 4.2 (1.4) | .39 | 20 (24.0) |
| *Entertainment* | 7 (11.47) | 353 (10.82) | 4.4 (1.2) | .04 | 14.9 (14.9) |
| *Multifeature* | 12 (19.67) | 1662 (50.96) | 4.4 (1.0) | <.001 | 25.8 (27.9) |

[a]Welch 2-sample *t* test between all reviews and each category.

[b]References to these categories are italicized in the text.

[c]N/A: not applicable.

## Word Clouds

Third, we report the word clouds of app reviews under each category in Multimedia Appendices 2-7. The word clouds represent the most common terms used in the app reviews. The more frequent a word, the bigger and more central its representation in the cloud. Generic words like *help* or *like/love* appeared across categories; also, category-specific words emerged, like *test* or *result* for *medical assessment* apps, *game* or *quotes* for *entertainment* apps, *people* or *chat* for *supportive resources* apps and *journal* or *meditation* for *therapeutic treatment* apps. However, some less predictable words also appeared and provided hints about the focus of the users. For instance, *severe* (ie, severe depression) appeared specifically at the top in the list of words for *medical assessment* apps or *inspire* in the list for *entertainment* apps, which included *quotes* apps.

## Sentiment Analysis of User Reviews

The LIWC sentiment analysis by feature is reported in Table 4 and Figure 2. The *P* value of the *t* test compares feature-specific reviews with other app reviews. For instance, the *analytical* score of 52.2 for *therapeutic treatment* versus an average score of 43 (SD 36.3) for all reviews has *P*<.001, meaning that it is significantly above average.

**Table 4.** Key sentiment dimensions by category[a].

| App category | Analytic rating | | Clout rating | | Authenticity rating | |
|---|---|---|---|---|---|---|
| | Value, mean (SD) | *P* value | Value, mean (SD) | *P* value | Value, mean (SD) | *P* value |
| *Psychoeducation* [b] | 46.7 (36.4) | .08 | 49.2 (33.4) | .001 | 34.8 (36.6) | <.001 |
| *Medical assessment* | 41 (38.1) | .18 | 29.5 (29.5) | <.001 | 48.9 (40.8) | .02 |
| *Therapeutic treatment* | 52.2 (36.3) | <.001 | 45.3 (31.2) | .15 | 36.8 (38.7) | <.001 |
| *Supportive resources* | 37.9 (35.0) | .08 | 47.6 (34.0) | .09 | 42.6 (40,2) | .43 |
| *Entertainment* | 31.8 (34.7) | <.001 | 47 (34.0) | .01 | 46.7 (40.0) | .50 |
| *Multifeature* | 44.2 (35.6) | .62 | 44.5 (32.5) | .01 | 47.15 (39.5) | .11 |
| All reviews | 43 (36.3) | N/A[c] | 42.8 (32.8) | N/A | 45.3 (39.7) | N/A |

[a]Welch 2-sample *t* test between all reviews and each category.

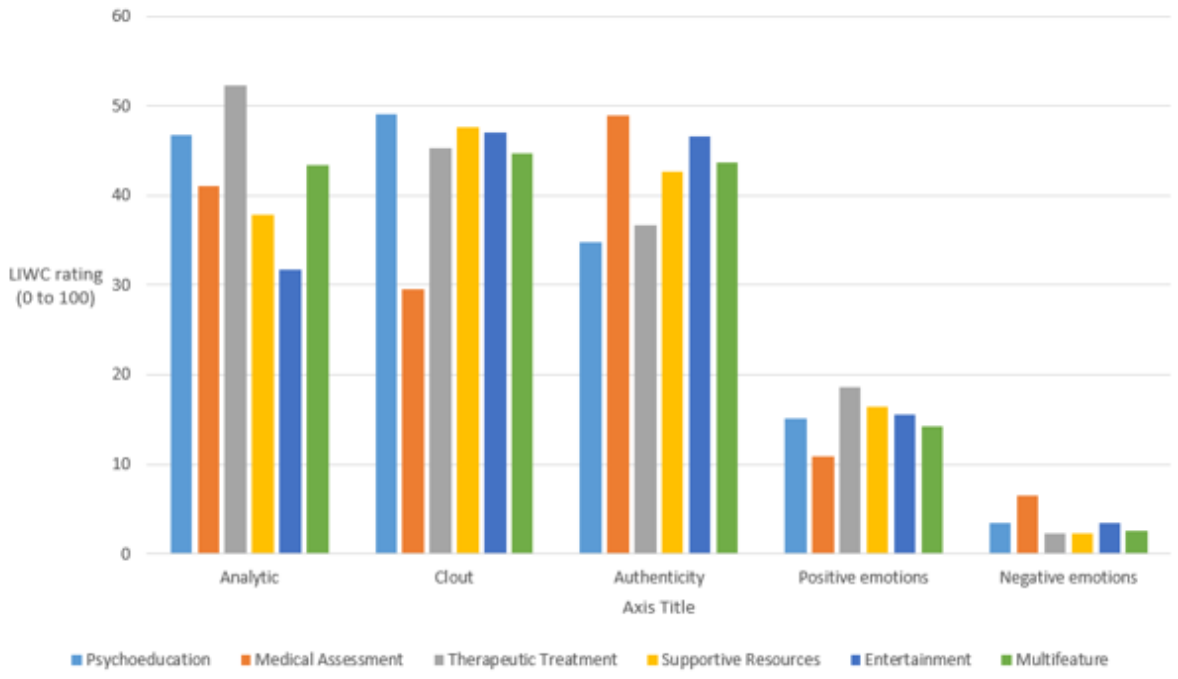
[b]References to these categories are italicized in the text.

[c]N/A: not applicable.



XSL•FO
RenderX

**Figure 2.** Key sentiment dimensions by single-feature app category. LIWC: Linguistic Inquiry Word Count.




*Psychoeducation* app reviews were significantly higher than the average on *clout*, but were less *authentic* (34.8), suggesting that the reviews were more focused on influencing others than on sharing personal experiences and that the users were more confident. This was illustrated by reviews such as the following:

> *No help at all. Lots of information that is easily available from a single Internet search.*
>
> *I found this helpful but it needs more information for another star, specially I felt that this app has no information about recovery & medical advice (verified doctors).*

*Medical assessment* app reviews had less *clout* (29.5) and more *authenticity* (48.9), but also significantly more negative emotions (6.6). This was illustrated by reviews such as, "Kinda lame ask general questions not really a benefit all ready knew answer" *or* "Only 20% of it was true about me in my test." Others leaned toward personal disclosure:

> *I got 25. I am 13 and i literally don't want my life anymore! I wonder how have i not suicided yet!*

*Therapeutic treatment* app reviews were more *analytical* (52.2) and less *authentic* (36.8), suggesting that the reviews were more focused on the actual functions. This was illustrated by reviews such as the following:

> *Brilliant app I particularly liked the progressive explanation of the cognitive distortions and how to address each accordingly.*
>
> *So helpful and easily accessible. You can pull the app out whenever you need it.*

*Supportive resources* app reviews exhibited less negative emotions (2.3) and anxiety (0.06) than the average. *Supportive resources* reviews focused on the process of connecting to other people as illustrated by, "This a great app to connect with people dealing with similar situations!" or on the app features, illustrated as follows:

> *I have been a member for a while now, and I love we got a mobile app. Wish we could take chats on it though. But awesome non the less.*

*Entertainment* app reviews exhibited feedback that went beyond entertainment concerns, such as the following*:*

> *This game is so amazing, I've been struggling with depression and self harm since I was 9, this game had such an emotional impact on me, I hope more people discover this game soon, yeah, its pretty laggy, but I think the over all message equals it out.*

Moreover, some were very negative, such as the following:

> *As a clinical counselor I would say this app is likely to lead suffers down a dark path.*

## Discussion

### Principal Findings

Findings pointed to differences in the emotional experiences of users based on the app features.

*Medical assessment* apps specifically received highly negative reviews. Their app ratings and positive emotions were significantly lower than the average of the depression app reviews, while the negative emotions were higher than the average. A possible explanation is that unlike other categories, *medical assessment* apps provide users with feedback and insights into their own personal conditions and whether they are depressed or not. Research suggests that people who disagree with personal feedback may respond with distress and exhibit strong and long-lasting feelings [37]. These emotions may translate into resentment and negative reactions against the quality and the validity of the app. The high level of authenticity



XSL•FO
RenderX

also suggests that the users reveal more about themselves. For developers, this suggests the need for caution before introducing *medical assessment* features in their apps, as they may antagonize their users and possibly distress them. It also raises the question of the impact of medical assessment on users. Distress caused by the outcome may lead the users to seek expert opinion for confirmation or disconfirmation, but it may also lead them to draw negative conclusions about what to expect from medical professionals.

In contrast, *supportive resources* apps generate fewer negative emotions. As these apps mostly connect the users to other people, the emotional response may focus more on the people or the services connected to them than on the apps themselves. *Therapeutic treatment* apps generate more positive emotions and anxiety but are also much more analytical, possibly because they focus on the users' attention on their actions to mitigate their condition. This suggests that both are safer features for developers to offer, at least to begin with. To a certain extent, *psychoeducation* apps also generate more positive reactions (although not significantly). Their low authenticity level can be explained by the impersonal informational and educational dimensions and it also makes them less risky to implement, to the extent that they do not mislead the users with incorrect information. Higher clout level is associated with higher confidence and social status. Users of psychoeducation apps may require sufficient self-confidence to believe that simply getting access to information is sufficient than more prescriptive features. Thus, these apps may cater to a different, more autonomous population.

*Entertainment* apps offer an ambivalent picture. They insignificantly generate more negative emotions, but their users also express significantly less anxiety. Entertainment could be an alternative way to engage people with depression who are anxious about dealing directly with their condition. This confirms findings from prior research obtained from focus groups, suggesting that people with mental health issues, especially male adolescents, value entertainment features in mental health management apps [15].

## Limitations

This study has limitations. The app marketplace is continuously evolving. The study was based solely on the information available in Google Play Store and Apple App Store. This information is subject to the inclusion criteria put in place by the app stores and the developers. The authenticity of the reviews included in this study was not validated, and the issue of illegitimate or fraudulent reviews is widespread [38]. These carry special risks regarding mental health apps, as they could lead the users to make decisions that may be detrimental to their health. User socio-demographic information was unavailable, even though this is a standard practice with recent studies involving sentiment analysis of app reviews [17,39,40].

Moreover, only the users who posted reviews of the apps were represented, which limits the sample's representativeness of the population of mental health app users. Reviewers were willing to publicly associate their usernames with a depression app, which many users may be reluctant to do, considering the stigma associated with mental disorders. In addition, even though the apps mostly cater to a population with depression, we do not know whether the reviewers are people diagnosed with clinical depression. Future research may try to replicate these findings by actively selecting respondents with depression and asking them to review apps. Such results can then be compared with those from this study or those from the app stores.

Finally, an assumption of this study is that the reviews can be generalized to the app feature, but they may reflect the idiosyncrasies of the reviewed apps (eg, bugs or ill-designed apps), as the number of apps in each category varies from 4 to 20.

## Comparison With Prior Work

This study contributes to the literature in multiple ways.

In addition to the studies that describe cognitive processes such as satisfaction or confirmation of expectations in mHealth app users [19], our findings suggest that depression apps also generate strong emotions. Emotions form a key element of mental health conditions and access to mental health care [3] and should be of concern to researchers and developers interested in improving the apps.

Depression app use is a health care behavior practiced by a large population with potentially serious mental health conditions [6]. The people routinely use depression apps to access information and assess, track, and manage their condition. Ultimately, they draw conclusions about their condition and take action (or maybe more problematically, do not take necessary action) based on feedback from these apps which could have critical impacts on their mental health if continued without clinical supervision. A major concern of researchers and clinicians regarding depression apps is the clinical validity of these interventions. Few of these apps are rigorously and clinically validated [5,12], and despite efforts to provide clinical evidence, the rapidly changing app environment and user behaviors do not suggest that use will be dominated by clinically validated apps in the near future [5]. Future studies could compare the user reviews of validated apps with those of nonvalidated apps.

Installations of depression apps vastly exceed the number of people accessing mental health care services, and therefore do not compete with traditional care as much as with not accessing care at all [6]. They are typically used as stand-alone self-help programs that are either poorly integrated or entirely not integrated with the continuum of care. How they fit in this continuum is a question by itself. Depression app features include clinically validated, inspired by sound research, alternative unproven approaches, or games with minimal or no clinical claims, several of which can be found in the same app. This indicates unclear differences between the apps used. As such, the use of mental health apps is of interest, both as a clinical intervention and as a common behavior performed by people with mental health conditions.

The reviews of entertainment apps, for instance, suggest that a nonclinical approach may provide help and relief to people with depression, which could lead them to acknowledge their condition, gain confidence in the value of external support, and seek other features in the apps. Thus, entertainment apps may



XSL•FO
RenderX

be a stepping stone that does not require the users to recognize their condition and their need for help, considering the stigma and emotional barriers associated with it. They could then serve as a gateway to recognizing the value and seeking professional care for people with serious mental health issues. In contrast, apps may act as a deterrent, either because the users feel that the apps are sufficient or even better than traditional care or because bad app experiences, such as an early and disturbing virtual depression assessment, would cause skepticism toward the value of medical expertise. This requires research on the pathways that patients follow between using depression apps and accessing traditional health care services. In line with approaches that follow up web-based health behaviors of specific groups of people [41], future research could longitudinally follow up depression app user behaviors and decision-making processes to identify these pathways. Such behavior patterns may help tailor apps to diverse populations.

Our findings also highlight the need to discriminate mHealth app use based on the features offered. Typically, researchers categorize apps according to the disease or condition addressed by the apps [6,7]. The features offered by the apps are important. Providing information about a condition and helping the users track their symptoms every day are very different services and our findings show that users react differently to them in their emotions and satisfaction. This suggests breaking down the study of mHealth apps based on the features they offer. This could come either in addition or as an alternative to studying mHealth apps based on the mental health condition that they address. Two mHealth apps that provide the same features for different diseases may have more in common than 2 apps that provide different features for the same disease. This study illustrates a novel way to investigate user beliefs and behaviors toward specific features. Although substantial efforts have been made to extract isolated reviews that specifically mention a feature [17-19], the size of the app ecosystem can make it possible to isolate apps that offer a single feature, thereby capturing reactions of the users who may not specifically mention the feature under consideration.

Sentiment analysis answers the need to rely on reactive vetting tools for mHealth apps (what Olff [5] refers to as *postmarketing surveillance*) in complement to randomized controlled trials. Researchers are increasingly recognizing the value of mining patient-generated web-based content and feedback [21] and this study is a step toward exploiting the potential of natural language content generated virtually by people with mental health disorders. Sentiment analysis of these data can help refine our understanding of how the users behave and react emotionally outside of clinical settings. How individuals communicate, what activities they engage in, and what language they use are potential indicators of mental health; users' mental health conditions, such as depression, may reveal patterns of web-based behaviors through Twitter feeds [16]. Further studies could assess the extent to which app reviewers fit into this pattern by comparing them with other app reviewers. Word clouds complement the insights provided by sentiment analysis. They provide an unmediated representation of the words and lexicons used by the reviewers. In addition to the emotions that are conveyed, we can see that the users focus on looking for *help* on whether they *like* or *lovegood* apps. We can deduce the typical focus of their reviews, such as *tests* and *results*, for *medical assessment*. They provide face value and a topical complement to sentiment analysis. Other language analysis tools such as Latent Dirichlet Allocation or Structural Topic Modelling could also be used to provide further insights into the app reviews.

One of the major appealing features of depression apps is that they allow people to circumvent the stigma associated with mental health issues and access services privately. Beyond the privacy-conscious population, our findings suggest that many users are willing to publicly share their personal and intimate experiences about depression on public outlets such as app stores. This source of data can be used to improve individual apps, understand general patterns of use, and learn about the beliefs, behaviors, and emotions of patients. This is also a cautious reminder that the users may not realize that they are not just talking to the community of depression app users but are making a public statement through both a personal and public Google or Apple account that can be viewed by the broader community, including people to whom the users may not want their condition to be revealed. Further research is needed to investigate the extent to which the app reviewers are aware of what they are disclosing and to whom.

## Conclusions

This study broadens our understanding of depression app use and user emotions and refines our knowledge of user experience based on the app features used. Users react with observably different emotions and sentiments depending on the features offered by the depression apps. This has implications for clinicians to better orient their patients to the proper apps and for developers to improve their design and handle the delicate and intimate aspects of a vulnerable population. It is also useful for users to better understand the risks and benefits of using mental health apps and for researchers to broaden their understanding of virtual behaviors of people with mental health.

Our understanding of the role of smartphones and other personal technologies, both as a cause of and as a solution to mental health disorders is still limited, and we need to broaden the scope of our investigations to include the emotions associated with these new behaviors in new and authentic data sources such as user reviews.

## Acknowledgments

The authors would like to thank the Ted Rogers School of Management for financing the project.



XSL•FO
RenderX

## Authors' Contributions

The first author (JM) designed the project, did the data collection and analysis, and wrote the paper. The second author (SO) wrote a first draft of the introduction and performed an initial data analysis on a subsample of the final data set.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Ranking and frequencies of the most frequent words by category of single-feature apps.
[PDF File (Adobe PDF File), 431 KB - formative_v5i12e17062_app1.pdf ]

Multimedia Appendix 2
Word cloud of psychoeducation app reviews.
[PNG File , 42 KB - formative_v5i12e17062_app2.png ]

Multimedia Appendix 3
Word cloud of medical assessment app reviews.
[PNG File , 38 KB - formative_v5i12e17062_app3.png ]

Multimedia Appendix 4
Word cloud of therapeutic treatment app reviews.
[PNG File , 25 KB - formative_v5i12e17062_app4.png ]

Multimedia Appendix 5
Word cloud of supportive resources app reviews.
[PNG File , 25 KB - formative_v5i12e17062_app5.png ]

Multimedia Appendix 6
Word cloud of entertainment app reviews.
[PNG File , 26 KB - formative_v5i12e17062_app6.png ]

Multimedia Appendix 7
Word cloud of multifeature app reviews.
[PNG File , 22 KB - formative_v5i12e17062_app7.png ]

## References

1. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. PLoS One 2015;10(2):1-14 [FREE Full text] [doi: 10.1371/journal.pone.0116820] [Medline: 25658103]
2. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
3. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. Ann Behav Med 2006 Dec;32(3):254-258. [doi: 10.1207/s15324796abm3203_12] [Medline: 17107299]
4. Marcus MA, Westra HA, Eastwood JD, Barnes KL, Mobilizing MR. What are young adults saying about mental health? An analysis of internet blogs. J Med Internet Res 2012 Jan;14(1):e17 [FREE Full text] [doi: 10.2196/jmir.1868] [Medline: 22569642]
5. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:1-8 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
6. Meyer J, Okuboyejo S. A functional analysis of depression apps user installs and rating. In: Proceedings of the Americas Conference on Information Systems. 2019 Presented at: Americas Conference on Information Systems; 2019; Cancun, Mexico URL: https://tinyurl.com/p9dm8jsk
7. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth 2015 Feb 16;3(1):e16 [FREE Full text] [doi: 10.2196/mhealth.3713] [Medline: 25689790]



XSL•FO
RenderX

8. Harrison V, Proudfoot J, Wee PP, Parker G, Pavlovic DH, Manicavasagar V. Mobile mental health: review of the emerging field and proof of concept study. J Ment Health 2011 Dec;20(6):509-524. [doi: 10.3109/09638237.2011.608746] [Medline: 21988230]
9. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
10. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013 Feb 07;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
11. Birney AJ, Gunn R, Russell JK, Ary DV. Moodhacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
12. Giosan C, Cobeanu O, Mogoaşe C, Szentagotai A, Mureşan V, Boian R. Reducing depressive symptomatology with a smartphone app: study protocol for a randomized, placebo-controlled trial. Trials 2017 May 12;18(1):215 [FREE Full text] [doi: 10.1186/s13063-017-1960-1] [Medline: 28494802]
13. Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. JMIR Mhealth Uhealth 2016 Aug 09;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
14. mHealth App Developer Economics Study 2015. Research 2 Guidance. 2016. URL: https://research2guidance.com/app-stores-the-number-one-distribution-channel-for-mhealth-apps/ [accessed 2021-11-10]
15. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: exploring adolescents' perspectives. Health Informatics J 2016 Jun;22(2):265-275 [FREE Full text] [doi: 10.1177/1460458214555041] [Medline: 25385165]
16. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June, 2014; Baltimore, Maryland, USA p. 51-60 URL: https://aclanthology.org/W14-3207.pdf
17. Guzman E, Maalej W. How do users like this feature? A fine grained sentiment analysis of app reviews. In: Proceedings of the IEEE 22nd International Requirements Engineering Conference (RE). 2014 Presented at: IEEE 22nd International Requirements Engineering Conference (RE); Aug. 25-29, 2014; Karlskrona, Sweden p. 153-162 URL: https://ieeexplore.ieee.org/abstract/document/6912257/
18. Keertipati S, Savarimuthu B, Licorish S. Approaches for prioritizing feature improvements extracted from app reviews. In: Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering. USA: ACM; 2016 Presented at: EASE '16: 20th International Conference on Evaluation and Assessment in Software Engineering; June 1 - 3, 2016; Limerick Ireland p. 1-6 URL: https://dl.acm.org/doi/abs/10.1145/2915970.2916003
19. Patel P, Licorish S, Savarimuthu B, MacDonell S. Studying expectation violations in socio-technical systems: a case study of the mobile app community. In: Proceedings of the 24th European Conference on Information Systems (ECIS, 2016). 2016 Presented at: 24th European Conference on Information Systems (ECIS, 2016); 2016; Istanbul, Turkey URL: https://aisel.aisnet.org/ecis2016_rp/14
20. Brucks M. A typology of consumer knowledge content. Adv Consum Res 1986;13:58-63 [FREE Full text]
21. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013 Nov 01;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
22. Huang H, Bashir M. Users' adoption of mental health apps: examining the impact of information cues. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e83 [FREE Full text] [doi: 10.2196/mhealth.6827] [Medline: 28659256]
23. Kim J, Gupta P. Emotional expressions in online user reviews: how they influence consumers' product evaluations. J Busin Res 2012 Jul;65(7):985-992. [doi: 10.1016/j.jbusres.2011.04.013]
24. Dadgar M, Joshi K. The role of information and communication technology in self-management of chronic diseases: an empirical investigation through value sensitive design. J Assoc Inform Syst 2018 Feb;19(2):86-112. [doi: 10.17705/jais1.00485]
25. de Guinea AO, Markus ML. Why break the habit of a lifetime? Rethinking the roles of intention, habit, and emotion in continuing information technology use. MIS Q 2009;33(3):433. [doi: 10.2307/20650303]
26. Abirami A, Askarunisa A. Sentiment analysis model to emphasize the impact of online reviews in healthcare industry. Online Inform Rev 2017 Aug 14;41(4):471-486. [doi: 10.1108/OIR-08-2015-0289]
27. Xiang Z, Du Q, Ma Y, Fan W. A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tourism Manag 2017 Feb;58:51-65. [doi: 10.1016/j.tourman.2016.10.001]
28. Maramba ID, Davey A, Elliott MN, Roberts M, Roland M, Brown F, et al. Web-based textual analysis of free-text patient experience comments from a survey in primary care. JMIR Med Inform 2015 May 06;3(2):e20 [FREE Full text] [doi: 10.2196/medinform.3783] [Medline: 25947632]
29. Hopper AM, Uriyo M. Using sentiment analysis to review patient satisfaction data located on the internet. J Health Organ Manag 2015;29(2):221-233 [FREE Full text] [doi: 10.1108/JHOM-12-2011-0129] [Medline: 25800334]




XSL•FO
RenderX

30. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
31. Ramirez-Esparza N, Chung CK, Kacewicz E, Pennebaker JW. The psychology of word use in depression forums in english and in Spanish: testing two text analytic approaches. In: Proceeedings of the International Conference on Web and Social Media. 2008 Presented at: International Conference on Web and Social Media; 2008; Seattle Washington USA p. 102-108 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/18623/18402
32. Pennebaker J, Boyd R, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. University of Texas at Austin. 2015. URL: https://repositories.lib.utexas.edu/handle/2152/31333 [accessed 2019-02-07]
33. Al-Natour S, Turetken O. A comparative assessment of sentiment analysis and star ratings for consumer reviews. Int J Inform Manag 2020 Oct;54:102132. [doi: 10.1016/j.ijinfomgt.2020.102132]
34. Evans AM, Stavrova O, Rosenbusch H. Expressions of doubt and trust in online user reviews. Comput Hum Behav 2021 Jan;114:106556. [doi: 10.1016/j.chb.2020.106556]
35. Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying words: predicting deception from linguistic styles. Pers Soc Psychol Bull 2003 May;29(5):665-675. [doi: 10.1177/0146167203029005010] [Medline: 15272998]
36. Hoon L, Vasa R, Schneider J, Grundy J. An analysis of the mobile app review landscape: trends and implications. Faculty of Information and Communication Technologies, Swinburne University of Technology. 2013. URL: https://researchbank.swinburne.edu.au/file/ba489b3c-a9bb-41f0-8b43-b6d1ff4d4829/1/PDF%20%2823%20pages%29.pdf [accessed 2021-11-10]
37. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. Adv Health Sci Educ Theory Pract 2008 Aug;13(3):275-288. [doi: 10.1007/s10459-006-9039-x] [Medline: 17091339]
38. Genc-Nayebi N, Abran A. A systematic literature review: opinion mining studies from mobile app store user reviews. J Syst Softw 2017 Mar;125:207-219. [doi: 10.1016/j.jss.2016.11.027]
39. Ranjan S, Mishra S. Comparative sentiment analysis of app reviews. In: Proceedings of the 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). 2020 Presented at: 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT); July 1-3, 2020; Kharagpur, India p. 1-7.
40. Liang T, Li X, Yang C, Wang M. What in consumer reviews affects the sales of mobile apps: a multifacet sentiment analysis approach. Int J Electron Comm 2015 Nov 30;20(2):236-260. [doi: 10.1080/10864415.2016.1087823]
41. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]

## Abbreviations

**LIWC:** Linguistic Inquiry Word Count
**mHealth:** mobile health

*Edited by G Eysenbach; submitted 14.11.19; peer-reviewed by M Rauws, D Leightley, B Chaudhry, Z Ma, T Muto, F Lanfranchi, C Fernández, T Ndabu; comments to author 14.09.20; revised version received 15.10.20; accepted 15.10.21; published 14.12.21.*

*Please cite as:*
*Meyer J, Okuboyejo S*
*User Reviews of Depression App Features: Sentiment Analysis*
*JMIR Form Res 2021;5(12):e17062*
*URL: https://formative.jmir.org/2021/12/e17062*
*doi:10.2196/17062*
*PMID:34904955*

©Julien Meyer, Senanu Okuboyejo. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Privacy and Confidentiality Concerns Related to the Use of mHealth Apps for HIV Prevention Efforts Among Malaysian Men Who Have Sex With Men: Cross-sectional Survey Study

Roman Shrestha[1], PhD; Celia Fisher[2], PhD; Jeffrey A Wickersham[3], PhD; Antoine Khati[1], MD; Rayne Kim[3], MPH; Iskandar Azwa[4], MD; Colleen Mistler[1], MPH; Lloyd Goldsamt[5], PhD

[1]Department of Allied Health Sciences, University of Connecticut, Storrs, CT, United States
[2]Center for Ethics Education, Fordham University, New York, NY, United States
[3]Department of Internal Medicine, Yale University, New Haven, CT, United States
[4]Department of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[5]Rory Meyers School of Nursing, New York University, New York, NY, United States

**Corresponding Author:**
Roman Shrestha, PhD
Department of Allied Health Sciences
University of Connecticut
358 Mansfield Road Unit 1101
Storrs, CT, 06269
United States
Phone: 1 860 486 2446
Email: roman.shrestha@uconn.edu

## Abstract

**Background:** The use of mobile health (mHealth), including smartphone apps, can improve the HIV prevention cascade for key populations such as men who have sex with men (MSM). In Malaysia, where stigma and discrimination toward MSM are high, the mHealth platform has the potential to open new frontiers for HIV prevention efforts. However, little guidance is available to inform researchers about privacy and confidentiality concerns unique to the development and implementation of app-based HIV prevention programs.

**Objective:** Given the lack of empirical data in this area, we aim to understand the privacy and confidentiality concerns associated with participation in a hypothetical app-based research study for HIV prevention efforts.

**Methods:** A cross-sectional, web-based survey was conducted between June and July 2020 among 355 Malaysian MSM. The survey included demographic and sexual health questions and a series of short videos describing a hypothetical app-based HIV prevention program, followed by questions related to privacy and confidentiality concerns in each step of the app-based program (ie, recruitment, clinical interaction, risk assessment, and weekly reminder). Multivariable logistic regression models were used to identify the correlates of willingness to use such an app-based program.

**Results:** Most of the participants (266/355, 74.9%) indicated their willingness to participate in a hypothetical mHealth app–based HIV prevention program. Participants expressed concerns about privacy, confidentiality, data security, and risks and benefits of participating in all stages of the app-based HIV research process. Multivariate analyses indicated that participants who had a higher degree of perceived participation benefits (adjusted odds ratio [aOR] 1.873; 95% CI 1.274-2.755; *P*=.001) were more willing to participate. In contrast, participants who had increased concerns about app-based clinical interaction and e-prescription (aOR 0.610; 95% CI 0.445-0.838; *P*=.002) and those who had a higher degree of perceived risks of participating (aOR 0.731; 95% CI 0.594-0.899; *P*=.003) were less willing to participate.

**Conclusions:** Overall, our results indicate that mHealth app–based HIV prevention programs are acceptable for future research on Malaysian MSM. The findings further highlighted the role of privacy and confidentiality, as well as the associated risks and benefits associated with participation in such a program. Given the ever-evolving nature of such technological platforms and the complex ethical–legal landscape, such platforms must be safe and secure to ensure widespread public trust and uptake.

***(JMIR Form Res 2021;5(12):e28311)*** doi:10.2196/28311



XSL•FO
RenderX

**KEYWORDS**

HIV; men who have sex with men; mHealth; ethics; Malaysia; mobile phone

## Introduction

### Background

HIV transmission among men who have sex with men (MSM) continues to accelerate and fuel the global epidemic, especially when stigma and discrimination are high [1,2]. Malaysia, a middle-income country with low (0.4%) overall HIV prevalence, has an evolving epidemic in MSM, with new HIV cases attributable to MSM increasing from 10% in 2008 to 54% in 2016 [3]. Biobehavioral surveys suggest that HIV prevalence among MSM is now at an all-time high of 21.6%. The highest prevalence is concentrated in the country's capital, Kuala Lumpur, where HIV prevalence among MSM was 43% in 2018, up from 22% in 2014 [4]. Central to the expanding HIV epidemic among Malaysian MSM is the concomitant use of drugs, especially amphetamine-type stimulants, linked to condomless sex and associated sexually transmitted infections [5-7]. In Malaysia, where both homosexuality and drug use are criminalized, MSM who use drugs bear the burden of social stigma and discrimination. Consequently, they are marginalized from traditional, venue-based HIV prevention approaches and require innovative HIV prevention strategies that improve coverage and access [8,9].

In the past decade, mobile phone technology has become increasingly ubiquitous and has allowed the use of new approaches, including mobile health (mHealth), to improve health outcomes [10-13]. mHealth tools can deliver effective HIV prevention in a confidential, less-stigmatizing, and convenient manner—all features that are crucial for prevention efforts among stigmatized and hard-to-reach populations, such as Malaysian MSM. MSM in Malaysia often do not access HIV prevention services, given the existing and limited service delivery models for the lesbian, gay, bisexual, and transgender (LGBT) community [14]. Recent studies with Malaysian MSM have shown that nearly all (>97%) of them own a smartphone. Many have stated preferences for interacting with smartphone apps to access HIV prevention services to avoid risk and shame because of disclosure [15,16]. As such, the use of HIV prevention apps holds promise for reducing transmission among this at-risk group [17-23], especially when linked to effective biomedical preventions, including pre-exposure prophylaxis [8,9,24]. However, no smartphone-based app tailored to the specific needs of MSM in Malaysia has been developed to improve access to HIV prevention services for this at-risk group.

Several studies have evaluated the efficacy of mHealth-based HIV prevention interventions targeting MSM populations globally [24-26]. mHealth is a trending research topic in Malaysia, as the number of published studies on the interest and use of mHealth technologies has exponentially increased over the past 10 years [27]. However, privacy and confidentiality risks remain to be primary concerns for sexual and gender minorities participating in mHealth research studies [28-35]. Privacy concerns around collecting sensitive information (eg, sexual and drug use behaviors) may limit willingness to use and engage in an HIV prevention mHealth app for MSM in Malaysia [28,36-38]. To date, no studies have examined whether any of these factors influence the decision among MSM to participate in an mHealth app–based HIV prevention program in Malaysia.

### Objectives

Understanding the motivations of MSM to participate in mHealth studies and the perceived privacy concerns is essential for increasing their representation in research designed to empirically validate prevention strategies. Given the lack of empirical data in this area of mHealth research and an existing mHealth app for HIV prevention in Malaysia, we aim to understand the perceived privacy and confidentiality issues associated with using an mHealth app for HIV prevention efforts among MSM in Malaysia. Specifically, this exploratory study evaluates the motivation of MSM in Malaysia to participate in a hypothetical mHealth app–based HIV prevention program. These data are crucial to developing an ethically sound, app-based HIV prevention program while minimizing any unintended harm, such as third-party use, incidental discovery by someone, or legal interception.

## Methods

### Study Design and Participants

A cross-sectional, web-based survey of Malaysian MSM was conducted from June to July 2020 to assess their attitudes toward participation in a hypothetical mHealth app–based HIV prevention program. Individuals were eligible if they (1) were ≥18 years; (2) identified as male; (3) had a self-reported HIV status as negative or unknown; (4) reported substance use or condomless anal sex with another man in the past 6 months; and (5) were able to read and understand English or Bahasa Malaysia.

### Study Procedures

A sample of MSM was recruited through purposive sampling via advertisements in 2 venues: geosocial networking (GSN) apps for MSM (eg, Grindr and Hornet) and popular social networking websites for the general population (eg, Facebook). The GSN apps pushed the advertisement as a message to the chat inboxes of all their different users in Malaysia. On the social networking site, we posted flyers on nongovernmental organizations and community-based organizations that provide services to MSM. In addition, we used banner advertisements that were targeted at MSM aged ≥18 years residing in Malaysia. These targeted advertisements could appear in one of two ways: a static advertisement on the right-hand pane of the website or an advertisement that resembled a standard post as users scrolled through their feeds. Interested users who clicked on advertisements were directed to an eligibility screening tool and a brief web-based consent form hosted by Qualtrics. The consent form indicated that they could "enter a lucky draw to win 1 of 10 RM 635 vouchers" (equivalent to US $150) and that there was "no participation necessary" to enter the random drawing.



XSL•FO
RenderX

Eligible volunteers completed a web-based consent form acknowledging that they understood the study's purpose, risks, and benefits before completing the survey. Those who declined to participate were provided instructions on how they could enter the drawing. Participants who completed the survey and were interested in entering the drawing were redirected to a separate website in which they entered an email address that was not linked to their data. On average, the participants took 10-12 minutes to complete the anonymous web-based survey. The study protocol and consent form were approved by the University of Malaya Research Ethics Committee and the Yale University Institutional Review Board.

We followed a protocol based on published standards for removing potentially duplicate cases while erring on the side of keeping rather than removing data in cases where a determination could not be made [39]. In particular, we identified potential duplicates based on age, sexual orientation, and ethnicity. All cases sharing those features in common were manually examined, focusing on responses to other questions, such as education, relationship status, income, device and browser information, and survey duration.

## Study Measures

### *Participant Characteristics*

These included age, sexual orientation, ethnicity, relationship status, educational attainment, income, self-reported history of HIV testing practices, disclosure of sexual orientation to others, and perceived stigma and discrimination from health care providers.

### *Smartphone Access and Use*

Participants were also asked about their access to a smartphone and use of the internet on a smartphone.

### *Sexual Risk Behaviors*

This included engagement, during the past 6 months, in anal intercourse with another man; engagement in condomless sex; number of sexual partners; and engagement in group sex, commercial sex, and chemsex. Chemsex is defined as the use of a psychoactive substance (eg, crystal methamphetamine or *batu* or ice, ketamine, ecstasy, and gamma-hydroxybutyrate *G*) before or during anal intercourse [40-42]. The perceived risk of acquiring HIV was measured by a single question, "What do you think is your current risk of getting HIV?" with a dichotomized response of "low" and "high."

### *Hypothetical mHealth App–Based HIV Prevention Program*

After completing the first part of the survey, participants were shown a short video that briefly introduced the hypothetical app-based HIV prevention program:

> *Imagine you have been asked to participate in a research study! In this study, you will be asked to use an app on your phone for HIV prevention. Advertisements would be placed on social networking apps like Grindr and Hornet to recruit participants into the program. After downloading the app, you will need to create an account, including creating a username and password and entering some personal information, like your name and date of birth. Once you're done, you can use this app for many things! For example, you can order pre-exposure prophylaxis (PrEP) through the app without having to visit a doctor in person. You can communicate with your doctor by sending messages and also receive reminders to take your medications. As part of the research study, you will also be asked to answer a few health-related questions.*

Next, participants were shown 4 additional videos illustrating each step involved in the hypothetical app-based program (eg, recruitment, clinical interaction, risk assessment, and weekly reminders), followed by 2-3 questions about ethical concerns (privacy, confidentiality, risks, and benefits) related to each step: for example, (1) recruitment—"Would you be concerned that others might find out about your sexual orientation if you click on the advertisement?"; (2) clinical interaction—"Would you be concerned that doctors and pharmacists receiving your medical information will share it with others?"; (3) risk assessment—"Would you be concerned that research staff might share the health-related information you provide in the app with other people?"; and (4) reminders—"Would you be concerned that other people might see the app messages/reminders on your phone?" In addition, 4 items each examined participants' overall perceived risks (eg, "I don't think all of my concerns will be addressed without seeing a doctor in person") and benefits (eg, "I could order PrEP through the app without having to visit a doctor in person") related to participating in a hypothetical app-based HIV prevention program. All responses to items within each domain (various steps of the hypothetical app program and perceived risks and benefits of participation) were summed to calculate a cumulative scale score with higher values indicating higher levels of concerns or benefits or risks.

## Data Analysis

Analyses were performed using SPSS (version 25.0; IBM Corp). Estimates were evaluated for statistical significance based on 95% CI using the probability criteria of $P<.05$. Descriptive statistics were calculated for all variables. The primary outcome was the willingness to use a smartphone app for HIV prevention. Next, we assessed whether there were demographic differences in the participation choice of MSM for a hypothetical app-based HIV prevention program (those willing vs those who were not willing). We conducted Pearson product–moment correlations between the likelihood that MSM would agree to participate in the hypothetical smartphone app and individual item and composite scale scores. Multivariate logistic regression was used to determine the extent to which cumulative scale scores and separate items together and independently accounted for variability in participation likelihood. The goodness-of-fit of the final multivariate model was assessed using the Hosmer–Lemeshow test [43].

# Results

During the 1-month recruitment period, 1259 participants entered the survey, and 47% (592/1259) consented and completed the screening tool. Of the 592 participants who



XSL•FO
RenderX

completed the screening and met the inclusion criteria, 227 (38.3%) did not complete the survey. There were no significant differences between completers and noncompleters for any demographic variables. Of the 365 (365/592, 61.6%) participants who completed the survey, 4 (1%) were eliminated because they failed validation checks (eg, survey duration) and 6 (1.6%) were not included in the analysis because they were identified as female, leaving a final sample of 355.

## Participant Characteristics

Multimedia Appendix 1 provides a summary of the characteristics of 355 participants with a mean age of 33.1 (SD 8.9) years. Most participants were single (238/355, 67%) and university graduates (300/355, 84.5%) and reported not disclosing their sexual orientation to family or friends (224/355, 63.1%). Overall, 22.8% (81/355) of participants had never been tested for HIV, and 42.5% (151/355) reported that they had been tested for HIV at least once in the past 6 months. In terms of sexual behaviors, 76.1% (270/355) of participants reported having anal sex with another man in the past 6 months, with 55.5% (150/270) reporting condomless sex. A minority (42/355, 11.8%) reported having engaged in chemsex, and 16.9% (60/355) of participants reported having engaged in group sex in the past 6 months. Of the 355 participants, 350 (98.6%) reported owning a smartphone and 274 (77.2%) indicated that they use a smartphone “all the time”. Notably, most participants (266/355, 74.9%) indicated their willingness to participate in the hypothetical mHealth app–based HIV prevention program.

## Privacy and Confidentiality Concerns of Using mHealth-Based HIV Prevention Apps

### *App-Based Recruitment*

Most participants indicated concerns related to GSN apps (eg, Grindr and Hornet) and their impact on their willingness to participate in mHealth research. Specifically, concerns about others finding out about the participants’ sexual orientation (235/255, 66.2%; $r=-0.166$; $P=.002$) and their HIV status (210/355, 59.2%; $r=-0.110$; $P=.04$) if they clicked on the advertisement were significantly associated with a reduced likelihood of participation. A composite score was created from the items in Table 1 listed under this category (Cronbach $\alpha=.771$; mean 1.9, SD 1.1), which was significantly associated with a reduced likelihood of participation ($r=-0.152$; $P=.004$).



XSL•FO
RenderX

**Table 1.** Survey item frequency and percent and correlation with the willingness to use a smartphone app for HIV prevention.

| Variables | Value, n (%) | Correlation |
|---|---|---|
| **Concerns related to app-based recruitment** | | |
| I'd be concerned that others might find out about my sexual orientation if I click on the advertisement. | 235 (66.2) | −0.166[a] |
| I'd be concerned that others might find out about my HIV status if I click on the advertisement. | 210 (59.2) | −0.110[b] |
| I'd be concerned that my online behavior might be tracked if I click on the advertisement. | 260 (73.2) | −0.100 |
| **Concerns related to app-based clinical interaction** | | |
| I'd be concerned that doctors and pharmacists receiving my medical information (eg, HIV test results) will share it with others. | 226 (63.7) | −0.167[a] |
| I'd be concerned about having medications delivered right to my home. | 147 (41.4) | −0.160[a] |
| I'd be worried because I wouldn't be seeing the doctor in person. | 167 (47) | −0.171[a] |
| **Concerns related to app-based risk assessment** | | |
| I'd be concerned about providing information about my sexual activities on the app. | 201 (56.6) | −0.152[a] |
| I'd be concerned about providing information about my drug use habit on the app. | 165 (46.5) | −0.126[b] |
| I'd be concerned that research staff might share the health-related information I provide in the app with other people. | 263 (74.1) | −0.120[b] |
| **Concerns related to receiving reminders** | | |
| I'd be concerned that other people might see the app messages/reminders on my phone. | 206 (58) | −0.150[a] |
| I'd be concerned about the frequency of the messages/reminders I receive on my phone. | 194 (54.6) | −0.070 |
| I'd be concerned about the timing of the messages/reminders I receive on my phone. | 192 (54.1) | −0.076 |
| **Benefits of participation** | | |
| I could order PrEP[c] through the app without having to visit a doctor in person. | 323 (91) | 0.068 |
| I would get reminder messages to take medications. | 323 (91) | 0.226[a] |
| I would have access to resources to learn more about HIV prevention. | 341 (96.1) | 0.117[b] |
| The results of the study could help other individuals like myself. | 345 (97.2) | 0.137[b] |
| **Risks of participation** | | |
| I don't think I would be comfortable using an app to get medications. | 101 (28.5) | −0.139[a] |
| I don't think all of my concerns will be addressed without seeing a doctor in person. | 164 (46.2) | −0.181[a] |
| I'm afraid about other people finding out about my participation in the study. | 191 (53.8) | −0.249[a] |
| I don't trust researchers to protect information I provide on the app. | 169 (47.6) | −0.268[a] |

[a]Correlation is significant at the significance level of 0.001.

[b]Correlation is significant at the significance level of 0.05.

[c]PrEP: pre-exposure prophylaxis.

### App-Based Clinical Interaction

A significant proportion (226/355, 63.7%) of the participants were concerned about clinical interaction via the app, particularly that the clinicians may share personal health information with others. All the items in this category (eg, clinicians sharing personal health information, home delivery of medications, and not being able to consult with doctors in person) were negatively correlated with participation choice, as was the composite score (Cronbach α=.509; mean 1.9, SD 1.5; $r$=−0.234; $P$<.001; Table 1).

### App-Based Risk Assessment

As illustrated in Table 1, items reflecting concerns related to using an app to assess HIV risk behaviors (sexual: 201/355, 56.6%, $r$=−0.152, $P$=.004; drug use: 165/355, 46.5%, $r$=−0.126, $P$=.02) were negatively correlated with the likelihood of participation in an app-based program. Interestingly, most (263/355, 74.1%) participants were concerned that the research staff might share the health-related information the participants provided via the app with other people outside the research study. The composite score (Cronbach α=.710; mean 1.7, SD 1.1) created with these items was significantly correlated with a reduced likelihood of participation ($r$=−0.167; $P$=.002).



XSL•FO
RenderX

### *Receiving Reminders*

Just over half of the MSM (206/255, 58%) indicated concerns about other people seeing the messages or reminders on their phones. This was significantly associated with a reduced likelihood of participation ($r$=−0.150; $P$=.005) as was the composite score for these items (Cronbach α=.876; mean 1.6, SD 1.3; $r$=−0.110; $P$=.38).

### Benefits of Participation

As shown in Table 1, nearly all participants endorsed the benefit of participating in the hypothetical mHealth app–based HIV prevention program. Of the 355 participants, 323 (90.9%) participants perceived receiving reminder messages to take medications, 341 (96.1%) participants perceived having an app-based platform to learn more about HIV prevention, and 345 (97.2%) participants perceived having altruistic motivation as benefits of participation, all of which were positively correlated with participation. The composite score constructed from these items (α=.620; mean 3.7, SD 0.6) also yielded a significant positive correlation with participation choice ($r$=0.198; $P$<.001).

### Risks of Participation

As shown in Table 1, participants reported various perceived risks associated with participation in such a hypothetical app-based study, all of which were negatively correlated with participation. Nearly one-fourth (74/355, 20.8%) of the participants indicated that they would not feel comfortable using an app to get HIV prevention medications. Almost half of the participants (164/355, 46.2%) indicated that their health concerns might not be addressed fully without seeing a health care provider in person. Similarly, 47.6% (169/355) of participants indicated that they did not trust researchers to protect the information they provided on the app. Furthermore, more than half of the participants (191/355, 53.8%) were concerned that other people might find out about their participation in the study. A composite score created from the items in Table 1 listed under this category (α=.748; mean 1.7, SD 1.4) was significantly associated with the likelihood of participation ($r$=−0.280; $P$<.001).

### Correlates of Willingness to Use a Smartphone App for HIV Prevention

Table 2 shows the independent correlates and aOR of willingness to use the smartphone app for HIV prevention. In the multivariate logistic regression model, those with a higher degree of perceived benefits (aOR 1.873; 95% CI 1.274-2.755; $P$=.001) were more willing to participate. Conversely, those with increased concerns related to app-based clinical interaction and e-prescription (aOR 0.610; 95% CI 0.445-0.838; 95% CI 0.594-0.899; $P$=.002) and those with a higher degree of perceived risks (aOR 0.731; $P$=.003) were less willing to participate.

**Table 2.** Multivariate logistic regression models of factors associated with willingness to use a smartphone app for HIV prevention.

| Variables | Willingness to use the app[a] | |
|---|---|---|
| | aOR[b] (95% CI) | *P* value |
| Age (years) | 0.973 (0.945-1.002) | .07 |
| **Highest education level** | | |
| Secondary education and below | Reference | __[c] |
| Tertiary education (college or university) | 1.617 (0.808-3.233) | .17 |
| **Concerns related to participation[d]** | | |
| App-based recruitment | 0.911 (0.682-1.218) | .53 |
| App-based clinical interaction | 0.610 (0.445-0.838) | .002 |
| App-based risk assessment | 0.844 (0.635-1.121) | .24 |
| Receiving reminders | 1.209 (0.947-1.544) | .13 |
| Benefits of participation[c] | 1.873 (1.274-2.755) | .001 |
| Risks of participation[c] | 0.731 (0.594-0.899) | .003 |

[a]$R^2$=0.239; Hosmer–Lemeshow test: $\chi^2_{8,355}$=8.0 $P$=.43.

[b]aOR: adjusted odds ratio.

[c]Reference group.

[d]Composite scores.

## Discussion

### Principal Findings

Our findings indicate that mHealth app–based HIV prevention programs are an acceptable research methodology for MSM in Malaysia. This finding is similar to those observed among MSM in other settings [44-49]. Our results also highlight the need to address ethical concerns related to privacy, confidentiality, and data security at all stages of the mHealth research process. These findings are significant in Malaysia, where homosexuality is considered a criminal behavior. Previous research has shown



XSL•FO
RenderX

that countries where homosexuality is criminalized often have higher rates of HIV among MSM and lower rates of access to HIV prevention and treatment services compared with their counterparts [50,51]. There are limited physical venues for MSM to meet face-to-face, socialize, and seek health care as homosexuality is illegal owing to the increasingly hostile sociopolitical environment facing LGBT people in Malaysia. With a large proportion of Malaysian MSM using smartphones to seek intimate partners and sexual health information [14], prevention programs need to use innovative platforms that increase the use of anonymous interfaces and reduce in-person interactions.

Although acceptable, our results indicated that Malaysian MSM were concerned about the potential data safety, privacy, and confidentiality concerns and risks and benefits associated with participating in a hypothetical app-based HIV prevention program. Specifically, concerns that information disclosed or shared via the app may be overheard or seen by unintended third parties (eg, colleagues or family members) were noted by more than half of the participants. Confidentiality concerns related to using an app for HIV prevention, including tracking app-based activity or sharing sensitive personal information (eg, health records or engagement in drug and sexual risk behaviors) disclosed via the app were also reported.

Despite these concerns, many participants indicated a willingness to participate, citing improved access to services without interacting with the clinicians in person, receiving personalized reminders, having greater access to information, and having a desire to contribute to a tradition of volunteerism in HIV prevention research [52-54]. These perceived benefits are significantly associated with willingness to participate and are particularly relevant in Malaysia. Previous research has demonstrated that stigma and discrimination are enacted on MSM by health care providers [55-57]. MSM are often hesitant to meet clinicians in person or disclose their sexuality and risk behaviors owing to fear of stigma, discrimination, or criminalization. This hesitation results in extraordinary health disparities and low HIV prevention uptake.

In addition, participants in our sample reported perceived risks associated with participation in mHealth research for HIV prevention, including an inability to fully address health concerns regarding the confidentiality of personal health information provided via the app. The perceived risks related to mHealth research participation negatively influenced willingness to participate. For example, an individual's perceived negative experiences within technological social spheres may impact their willingness to engage in an mHealth platform, especially for those who feel their sexual orientation or substance use and sexual history could be accessed by, or disclosed without consent to, law enforcement [58,59]. Addressing potential confidentiality, privacy, and data management concerns and relaying these safeguards to participants could increase their willingness to participate. With the increasing use of app-based platforms for HIV prevention services globally, current findings indicate the need to assess participants' preferences for and feedback to improve engagement in such technological platforms.

## Implications

The findings of this study have significant implications for developing and implementing mHealth app–based programs among MSM in low- and middle-income countries. There is value in using an mHealth app among Malaysian MSM to transform the face of HIV prevention service delivery and personal health management. Leveraging mHealth app–based programs reduces individuals' discomfort and distrust of disclosing risk behaviors and clinicians' low cultural competency to work with individuals of diverse sexual identities. Doing so reduces barriers to accessing health care for marginalized populations, such as MSM [60-64], who would have otherwise not opted to seek care. Such an app-based platform would help address issues related to accessibility and use of HIV prevention programs among hard-to-reach and marginalized groups.

The findings highlight many privacy and confidentiality concerns specific to mHealth platforms that need to be addressed to realize their full potential. These concerns become especially relevant in Malaysia and other regions of the world where the burden of social stigma and discrimination is high and where drug use and homosexual behavior are punishable by law. In particular, addressing potential confidentiality and data management concerns before implementing an app-based program and relaying appropriate safeguards to participants are crucial to facilitate widespread public trust and implementation. Such concerns could be addressed by controlling data storage and transmission and accounting for potential third-party data access [27]. The risk of privacy violations can also be minimized by considering how the data are stored locally on the device or in the cloud or transmitted to the research or clinical team. Apps can be placed in secure vaults on mobile devices or can be entirely password-protected with potential secondary authentication measures (eg, biometrics or personal ID number verification) [27,52].

Participants should be informed of the benefits and risks of participating in mHealth research—such as third-party access to data collected via hacking, legal interception (eg, a subpoena), and incidental discovery by someone accessing the phone [27,56]. The developers should maximize opportunities for individuals to control what data they share with the clinical or research team, such as using pop-up messages asking whether they consent to the specific information being transmitted. In addition, controlling the timing of reminders, their form of transmission (such as SMS text messaging vs pop-up message), and coding of messages in a way that only the intended recipient would understand are other ways to prevent a breach in confidentiality. This approach can be tailored according to personal preferences and the participant's comfort in disclosing health-related information. Furthermore, ensuring data confidentiality through layers of encryption and establishing a secure network connection can prevent remote access to information by potential hackers [52]. Overall, as many of this study's findings apply to mHealth research among MSM, this study may inform the development and implementation of mHealth app–based HIV prevention programs among key populations in Malaysia and other low- and middle-income countries.



XSL•FO
RenderX

## Limitations

This study had several limitations. First, we used a convenience sample of MSM recruited via gay social media (eg, Grindr and Hornet) rather than a representative sample of all MSM in Malaysia. The web-based methodology in this study may have excluded MSM without access to the internet or those who did not use social networking platforms (eg, Grindr, Hornet, or Facebook) on which the study was advertised. Furthermore, a significant proportion of participants in this study were younger and well-educated, limiting the generalizability of our findings. Second, the use of self-reported measures in this study may have resulted in participants inaccurately reporting socially undesirable behaviors. Third, the use of a cross-sectional study design limited our ability to make causal inferences. Finally, the findings from this study revolved around the features of a hypothetical mHealth app rather than any existing app. Therefore, future research including both MSM (inclusive of both young and older individuals and across the spectrum of education status) and clinical stakeholders is critical. Such research should iteratively solicit feedback at multiple stages of app development to improve acceptability and user experience.

## Conclusions

This study found a high willingness to use a hypothetical app-based HIV prevention program among MSM in Malaysia. The findings further highlighted the role of privacy, confidentiality, and risks and benefits associated with participation in such a program. These findings indicate that privacy and confidentiality concerns related to using an app for HIV prevention efforts are real and critical to ensure widespread public trust and uptake. Therefore, it is important that such platforms are safe and secure. Given the ever-evolving nature of such technological solutions and the complex ethical–legal landscape, the development and implementation of a scientifically and ethically sound app-based HIV prevention program will require early and ongoing engagement with MSM and other relevant stakeholders.

## Acknowledgments

RS was supported in part by a career development award from the National Institute on Drug Abuse (K01 DA051346; principal investigator [PI]: RS) and a research grant from the Fogarty International Center (R21 TW011665; PI: RS). Data collection for this study was supported by the Fordham HIV Prevention Research Ethics Training Institute via a training grant sponsored by the National Institute on Drug Abuse (R25-DA031608; PI: CBF). The funders had no role in the study design; data collection; and analysis, decision to publish, or preparation of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant characteristics of Malaysian men who have sex with men, stratified by willingness to use a smartphone app for HIV prevention.
[DOCX File , 26 KB - formative_v5i12e28311_app1.docx ]

## References

1. Global HIV statistics. UNAIDS. 2017. URL: https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf [accessed 2019-05-01]
2. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul;380(9839):367-377. [doi: 10.1016/s0140-6736(12)60821-6]
3. Global AIDS Response Progress Report: Malaysia 2016. Ministry of Health, Malaysia. 2016. URL: https://www.unaids.org/sites/default/files/MYS_narrative_report_2016.pdf [accessed 2021-12-06]
4. Integrated Biological and Behavioral Surveillance Survey 2017. Ministry of Health Malaysia. 2019. URL: https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/Laporan_Kajian_IBBS_2017.pdf [accessed 2021-12-14]
5. Lim SH, Bazazi AR, Sim C, Choo M, Altice FL, Kamarulzaman A. High rates of unprotected anal intercourse with regular and casual partners and associated risk factors in a sample of ethnic Malay men who have sex with men (MSM) in Penang, Malaysia. Sex Transm Infect 2013 Dec 20;89(8):642-649. [doi: 10.1136/sextrans-2012-050995] [Medline: 23787168]
6. Kanter J, Koh C, Razali K, Tai R, Izenberg J, Rajan L, et al. Risk behaviour and HIV prevalence among men who have sex with men in a multiethnic society: a venue-based study in Kuala Lumpur, Malaysia. Int J STD AIDS 2011 Jan 01;22(1):30-37. [doi: 10.1258/ijsa.2010.010277] [Medline: 21364064]
7. Lim SH, Cheung DH, Guadamuz TE, Wei C, Koe S, Altice FL. Latent class analysis of substance use among men who have sex with men in Malaysia: findings from the Asian Internet MSM Sex Survey. Drug Alcoh Depend 2015 Jun;151:31-37. [doi: 10.1016/j.drugalcdep.2015.02.040]
8. United Nations Development Programme. The policy and legal environments related to HIV services in Malaysia: review and consultation. United Nations, Malaysia. 2014. URL: https://hivlawcommission.org/wp-content/uploads/2017/06/rbap-hhd-2014-policy-and-legal-envt-related-to-hiv-services-in-malaysia-1.pdf [accessed 2021-12-06]



XSL•FO
RenderX

9. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
10. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth interventions to support self-management in HIV: a systematic review. Open AIDS J 2017 Nov 21;11(1):119-132 [FREE Full text] [doi: 10.2174/1874613601711010119] [Medline: 29290888]
11. Henny KD, Wilkes AL, McDonald CM, Denson DJ, Neumann MS. A rapid review of eHealth interventions addressing the continuum of HIV care (2007-2017). AIDS Behav 2018 Jan 5;22(1):43-63 [FREE Full text] [doi: 10.1007/s10461-017-1923-2] [Medline: 28983684]
12. Conserve DF, Jennings L, Aguiar C, Shin G, Handler L, Maman S. Systematic review of mobile health behavioural interventions to improve uptake of HIV testing for vulnerable and key populations. J Telemed Telecare 2016 Jul 09;23(2):347-359. [doi: 10.1177/1357633x16639186]
13. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873]
14. Shrestha R, Lim S, Altice F, Copenhaver M, Saifi R, Ab Halim MA, et al. Use of smartphone to find sexual health information online among Malaysian MSM: implications for mHealth intervention. Annals of Behavioral Medicine. 2019. URL: https://tinyurl.com/mry9zhx7 [accessed 2021-12-06]
15. Shrestha R, Lim SH, Altice FL, Kamarulzaman A, Wickersham JA. Perception and acceptability of mHealth-based intervention for HIV prevention among men who have sex with men (MSM) in Malaysia. In Preparation 2019.
16. Shrestha R, Lim SH, Altice FL, Copenhaver M, Wickersham JA, Saifi R, et al. Use of smartphone to seek sexual health information online among Malaysian men who have sex with men (MSM): implications for mHealth intervention to increase HIV testing and reduce HIV risks. J Community Health 2020 Feb 2;45(1):10-19 [FREE Full text] [doi: 10.1007/s10900-019-00713-x] [Medline: 31375976]
17. Chawarski MC, Vicknasingam B, Mazlan M, Schottenfeld RS. Lifetime ATS use and increased HIV risk among not-in-treatment opiate injectors in Malaysia. Drug Alcohol Depend 2012 Jul 01;124(1-2):177-180 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.12.024] [Medline: 22266088]
18. Fendrich M, Avci O, Johnson TP, Mackesy-Amiti ME. Depression, substance use and HIV risk in a probability sample of men who have sex with men. Addict Behav 2013 Mar;38(3):1715-1718 [FREE Full text] [doi: 10.1016/j.addbeh.2012.09.005] [Medline: 23254224]
19. Daskalopoulou M, Rodger A, Phillips AN, Sherr L, Speakman A, Collins S, et al. Recreational drug use, polydrug use, and sexual behaviour in HIV-diagnosed men who have sex with men in the UK: results from the cross-sectional ASTRA study. Lancet HIV 2014 Oct;1(1):22-31. [doi: 10.1016/s2352-3018(14)70001-3]
20. Hirshfield S, Remien RH, Humberstone M, Walavalkar I, Chiasson MA. Substance use and high-risk sex among men who have sex with men: a national online study in the USA. AIDS Care 2004 Nov 27;16(8):1036-1047. [doi: 10.1080/09540120412331292525] [Medline: 15511735]
21. Van Tieu H, Koblin BA. HIV, alcohol, and noninjection drug use. Curr Opin HIV AIDS 2009;4(4):314-318. [doi: 10.1097/coh.0b013e32832aa902]
22. Halkitis PN, Parsons JT. Recreational drug use and HIV-risk sexual behavior among men frequenting gay social venues. J Gay Lesbian Soc Serv 2008 Sep 22;14(4):19-38. [doi: 10.1300/j041v14n04_02]
23. Sandfort T, Knox J, Alcala C, El-Bassel N, Kuo I, Smith L. Substance use and HIV risk among men who have sex with men in Africa: a systematic review. J Acquir Immune Defic Syndr 2017 Oct 01;76(2):34-46 [FREE Full text] [doi: 10.1097/QAI.0000000000001462] [Medline: 28903126]
24. Moore D, Jain S, Dubé MP, Daar ES, Sun X, Young J, et al. Randomized controlled trial of daily text messages to support adherence to preexposure prophylaxis in individuals at risk for human immunodeficiency virus: the TAPIR Study. Clin Infect Dis 2018 May 02;66(10):1566-1572 [FREE Full text] [doi: 10.1093/cid/cix1055] [Medline: 29228144]
25. Muessig K, LeGrand S, Horvath K, Bauermeister J, Hightow-Weidman L. Recent mobile health interventions to support medication adherence among HIV-positive MSM. Curr Opin HIV AIDS 2017 Sep;12(5):432-441 [FREE Full text] [doi: 10.1097/COH.0000000000000401] [Medline: 28639990]
26. Liu A, Vittinghoff E, von Felten P, Amico KR, Anderson PL, Lester R, et al. Randomized controlled trial of a mobile health intervention to promote retention and adherence to preexposure prophylaxis among young people at risk for human immunodeficiency virus: the EPIC Study. Clin Infect Dis 2019 May 30;68(12):2010-2017 [FREE Full text] [doi: 10.1093/cid/ciy810] [Medline: 30239620]
27. Khan N, Qureshi M, Mustapha I, Irum S, Arshad R. A systematic literature review paper on online medical mobile applications in Malaysia. Int J Onl Eng 2020 Jan 21;16(01):63. [doi: 10.3991/ijoe.v16i01.12263]
28. Carter A, Liddle J, Hall W, Chenery H. Mobile phones in research and treatment: ethical guidelines and future directions. JMIR Mhealth Uhealth 2015 Oct 16;3(4):e95 [FREE Full text] [doi: 10.2196/mhealth.4538] [Medline: 26474545]
29. Labrique AB, Kirk GD, Westergaard RP, Merritt MW. Ethical issues in mHealth research involving persons living with HIV/AIDS and substance abuse. AIDS Res Treat 2013;2013:189645-189646 [FREE Full text] [doi: 10.1155/2013/189645] [Medline: 24171110]




XSL•FO
RenderX

30. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of mobile health in communities underrepresented in biomedical research: barriers and ethical considerations for scientists. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e87 [FREE Full text] [doi: 10.2196/mhealth.6494] [Medline: 28659258]
31. Hall JL, McGraw D. For telehealth to succeed, privacy and security risks must be identified and addressed. Health Aff (Millwood) 2014 Feb;33(2):216-221. [doi: 10.1377/hlthaff.2013.0997] [Medline: 24493763]
32. Bonar EE, Koocher GP, Benoit MF, Collins RL, Cranford JA, Walton MA. Perceived risks and benefits in a text message study of substance abuse and sexual behavior. Ethics Behav 2018 Mar 10;28(3):218-234 [FREE Full text] [doi: 10.1080/10508422.2017.1293541] [Medline: 29632430]
33. Broaddus MR, Marsch LA. Risks and benefits of text message-delivered and small group-delivered sexual health interventions among African American women in the Midwestern U.S. Ethics Behav 2015 Mar;25(2):146-168 [FREE Full text] [doi: 10.1080/10508422.2014.948957] [Medline: 25750497]
34. Burden K, Schuck S, Aubusson P. m-Research: ethical issues in researching young people's use of mobile devices. Youth Studies Australia. 2012. URL: https://childethics.com/library/journal-articles/m-research-ethical-issues-in-researching-young-peoples-use-of-mobile-devices/ [accessed 2021-12-06]
35. Khosropour CM, Sullivan PS. Risk of disclosure of participating in an internet-based HIV behavioural risk study of men who have sex with men. J Med Ethics 2011 Jun 27;37(12):768-769. [doi: 10.1136/jme.2011.043976]
36. Broaddus MR, DiFranceisco WJ, Kelly JA, St Lawrence JS, Amirkhanian YA, Dickson-Gomez JD. Social media use and high-risk sexual behavior among black men who have sex with men: a three-city study. AIDS Behav 2015 Jun 8;19 Suppl 2(S2):90-97 [FREE Full text] [doi: 10.1007/s10461-014-0980-z] [Medline: 25566762]
37. Yang C, Linas B, Kirk G, Bollinger R, Chang L, Chander G, et al. Feasibility and acceptability of smartphone-based ecological momentary assessment of alcohol use among African American men who have sex with men in Baltimore. JMIR Mhealth Uhealth 2015 Jun 17;3(2):e67 [FREE Full text] [doi: 10.2196/mhealth.4344] [Medline: 26085078]
38. Garett R, Menacho L, Young SD. Ethical issues in using social media to deliver an HIV prevention intervention: results from the HOPE Peru Study. Prev Sci 2017 Feb 9;18(2):225-232 [FREE Full text] [doi: 10.1007/s11121-016-0739-z] [Medline: 27933425]
39. Teitcher JE, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics 2015 Jan 01;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
40. Schmidt AJ, Bourne A, Weatherburn P, Reid D, Marcus U, Hickson F, EMIS Network. Illicit drug use among gay and bisexual men in 44 cities: findings from the European MSM Internet Survey (EMIS). Int J Drug Policy 2016 Dec;38:4-12. [doi: 10.1016/j.drugpo.2016.09.007] [Medline: 27788450]
41. Daly M. The meth-fuelled, wee-long orgies ravaging London's gay sex party scene. Vice. 2013. URL: https://www.vice.com/en/article/9bnxnv/the-week-long-meth-fuelled-sex-parties-taking-over-londons-gay-scene [accessed 2021-12-06]
42. Stephenson G, Richardson A. New psychoactive substances in England: a review of the evidence. GOV.UK Blog. 2014. URL: https://www.gov.uk/government/publications/new-psychoactive-substances-in-england-a-review-of-the-evidence [accessed 2021-12-06]
43. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997 May 15;16(9):965-980. [doi: 10.1002/(sici)1097-0258(19970515)16:9<965::aid-sim509>3.0.co;2-o]
44. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of smartphone application-based HIV prevention among young men who have sex with men. AIDS Behav 2014 Feb 1;18(2):285-296 [FREE Full text] [doi: 10.1007/s10461-013-0671-1] [Medline: 24292281]
45. Grosskopf NA, LeVasseur MT, Glaser DB. Use of the internet and mobile-based "apps" for sex-seeking among men who have sex with men in New York City. Am J Mens Health 2014 Nov 20;8(6):510-520 [FREE Full text] [doi: 10.1177/1557988314527311] [Medline: 24658284]
46. Levy ME, Watson CC, Wilton L, Criss V, Kuo I, Glick SN, et al. Acceptability of a mobile smartphone application intervention to improve access to HIV prevention and care services for black men who have sex with men in the district of Columbia. Digit Cult Educ 2015;7(2):169-191 [FREE Full text] [Medline: 26594251]
47. Zhang A, Reynolds NR, Farley JE, Wang X, Tan S, Yan J. Preferences for an HIV prevention mobile phone app: a qualitative study among men who have sex with men in China. BMC Public Health 2019 Mar 12;19(1):297 [FREE Full text] [doi: 10.1186/s12889-019-6617-4] [Medline: 30866880]
48. Sullivan PS, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and acceptability of a mobile comprehensive hiv prevention app for men who have sex with men: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26 [FREE Full text] [doi: 10.2196/mhealth.7199] [Medline: 28279949]
49. Goedel WC, Mitchell JW, Krebs P, Duncan DT. Willingness to use mobile phone apps for HIV prevention among men who have sex with men in London: web-based survey. JMIR Mhealth Uhealth 2017 Oct 11;5(10):e153 [FREE Full text] [doi: 10.2196/mhealth.8143] [Medline: 29021132]
50. Lyons C. Utilizing individual level data to assess the relationship between prevalent HIV infection and punitive same sex policies and legal barriers across 10 countries in Sub-Saharan Africa. In: Proceedings of the 23rd Virtual International



XSL•FO
RenderX

AIDS Conference. 2020 Presented at: 23rd Virtual International AIDS Conference; July 6-10, 2020; Virtual URL: https://www.aids2020.org/wp-content/uploads/2020/09/AIDS2020_Abstracts.pdf
51. Evaluation of the global commission on HIV and the law. Global Commission on HIV and the Law. 2021. URL: https://hivlawcommission.org/evaluation-of-the-global-commission-on-hiv-the-law/ [accessed 2021-12-06]
52. Pasipanodya EC, Kohli M, Fisher CB, Moore DJ, Curtis B. Perceived risks and amelioration of harm in research using mobile technology to support antiretroviral therapy adherence in the context of methamphetamine use: a focus group study among minorities living with HIV. Harm Reduct J 2020 Jun 11;17(1):41. [doi: 10.1186/s12954-020-00384-1] [Medline: 32527276]
53. Balfour L, Corace K, Tasca GA, Tremblay C, Routy J, Angel JB. Altruism motivates participation in a therapeutic HIV vaccine trial (CTN 173). AIDS Care 2010 Oct 09;22(11):1403-1409. [doi: 10.1080/09540121003693522]
54. Henderson GE, Peay HL, Kroon E, Cadigan RJ, Meagher K, Jupimai T, et al. Ethics of treatment interruption trials in HIV cure research: addressing the conundrum of risk/benefit assessment. J Med Ethics 2018 Apr 10;44(4):270-276 [FREE Full text] [doi: 10.1136/medethics-2017-104433] [Medline: 29127137]
55. Earnshaw VA, Jin H, Wickersham J, Kamarulzaman A, John J, Altice FL. Exploring intentions to discriminate against patients living with HIV/AIDS among future healthcare providers in Malaysia. Trop Med Int Health 2014 Jun 26;19(6):672-679 [FREE Full text] [doi: 10.1111/tmi.12306] [Medline: 24666546]
56. Earnshaw VA, Jin H, Wickersham JA, Kamarulzaman A, John J, Lim SH, et al. Stigma toward men who have sex with men among future healthcare providers in Malaysia: would more interpersonal contact reduce prejudice? AIDS Behav 2016 Jan 1;20(1):98-106 [FREE Full text] [doi: 10.1007/s10461-015-1168-x] [Medline: 26324078]
57. Jin H, Earnshaw VA, Wickersham JA, Kamarulzaman A, Desai MM, John J, et al. An assessment of health-care students' attitudes toward patients with or at high risk for HIV: implications for education and cultural competency. AIDS Care 2014 Mar 14;26(10):1223-1228 [FREE Full text] [doi: 10.1080/09540121.2014.894616] [Medline: 24625279]
58. Rudolph AE, Young AM, Havens JR. A rural/urban comparison of privacy and confidentiality concerns associated with providing sensitive location information in epidemiologic research involving persons who use drugs. Addict Behav 2017 Nov;74:106-111 [FREE Full text] [doi: 10.1016/j.addbeh.2017.06.006] [Medline: 28609723]
59. Hubach RD, O'Neil A, Stowe M, Giano Z, Curtis B, Fisher CB. Perceived confidentiality risks of mobile technology-based ecologic momentary assessment to assess high-risk behaviors among rural men who have sex with men. Arch Sex Behav 2021 May 20;50(4):1641-1650. [doi: 10.1007/s10508-019-01612-x] [Medline: 32078710]
60. WHO Global Observatory for eHealth. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva: World Health Organization; 2011.
61. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan 15;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
62. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov;376(9755):1838-1845. [doi: 10.1016/s0140-6736(10)61997-6]
63. Milward J, Lynskey M, Strang J. Solving the problem of non-attendance in substance abuse services. Drug Alcohol Rev 2014 Nov 06;33(6):625-636. [doi: 10.1111/dar.12194] [Medline: 25196817]
64. Mbuagbaw L, Mursleen S, Lytvyn L, Smieja M, Dolovich L, Thabane L. Mobile phone text messaging interventions for HIV and other chronic diseases: an overview of systematic reviews and framework for evidence transfer. BMC Health Serv Res 2015 Jan 22;15(1):33 [FREE Full text] [doi: 10.1186/s12913-014-0654-6] [Medline: 25609559]

## Abbreviations

**aOR:** adjusted odds ratio
**CS:** condomless sex
**GSN:** geosocial networking
**LGBT:** lesbian, gay, bisexual, and transgender
**MSM:** men who have sex with men
**PI:** principal investigator



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 05.03.21; peer-reviewed by R Hubach, J Stekler; comments to author 25.06.21; revised version received 19.07.21; accepted 21.11.21; published 16.12.21.*

*Please cite as:*
*Shrestha R, Fisher C, Wickersham JA, Khati A, Kim R, Azwa I, Mistler C, Goldsamt L*
*Privacy and Confidentiality Concerns Related to the Use of mHealth Apps for HIV Prevention Efforts Among Malaysian Men Who Have Sex With Men: Cross-sectional Survey Study*
*JMIR Form Res 2021;5(12):e28311*
*URL: https://formative.jmir.org/2021/12/e28311*
*doi:10.2196/28311*
*PMID:34924355*

©Roman Shrestha, Celia Fisher, Jeffrey A Wickersham, Antoine Khati, Rayne Kim, Iskandar Azwa, Colleen Mistler, Lloyd Goldsamt. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# The Impact of Patient Characteristics on Their Attitudes Toward an Online Patient Portal for Communicating Laboratory Test Results: Real-World Study

Rosian Tossaint-Schoenmakers[1,2,3], MSc; Marise Kasteleyn[2,3], MSc, PhD; Annelijn Goedhart[1], BA; Anke Versluis[2,3], MSc, PhD; Esther Talboom-Kamp[2,3,4], MBA, MD, PhD

[1]Saltro Diagnostic Centre, Utrecht, Netherlands
[2]National eHealth Living Lab, Leiden University Medical Centre, Leiden, Netherlands
[3]Public Health and Primary Care Department, Leiden University Medical Centre, Leiden, Netherlands
[4]Unilabs Group, Geneva, Switzerland

**Corresponding Author:**
Rosian Tossaint-Schoenmakers, MSc
Saltro Diagnostic Centre
Mississippidreef 83
Utrecht, 3565 CE
Netherlands
Phone: 31 302361136
Email: rtossaint@saltro.nl

## Abstract

**Background:** Patient portals are promising tools to increase patient involvement and allow them to manage their health. To optimally facilitate patients, laboratory test results should be explained in easy language. Patient characteristics affect the usage of portals and the user satisfaction. However, limited research is available, specified for online communicating laboratory test results, on whether portal use and acceptance differ between groups.

**Objective:** The aim of this study was to assess the effect of patient characteristics (gender, age, education, and chronic disease) on the self-efficacy and perceived usability of an online patient portal that communicates diagnostic test results.

**Methods:** We used the online-administered eHealth impact questionnaire (eHIQ) to explore patients' attitudes toward the portal. Patients visiting the portal were asked to complete the questionnaire and to answer questions regarding gender, age, education, and chronic disease. The subscale "information and presentation" of the eHIQ assessed the usability of the patient portal and the subscale "motivation and confidence to act" assessed self-efficacy to determine whether patients were motivated to act on the presented information. Age, gender, education, and chronic disease were the determinants to analyze the effect on usability and self-efficacy. Descriptive analyses were performed to explore patient characteristics, usability, and self-efficacy. Univariable and multivariable regression analyses were performed with age, gender, education, and chronic disease as determinants, and usability and self-efficacy as outcomes.

**Results:** The questionnaire was completed by 748 respondents, of which 428 (57.2%) were female, 423 (56.6%) were highly educated, and 509 (68%) had no chronic disease. The mean age was 58.5 years (SD 16.4). Higher age, high education, and asthma or chronic obstructive pulmonary disease were significant determinants for decreased usability; respectively, b=-.094, 95% CI -1147 to 0.042 ($P$<.001); b=-2.512, 95% CI -4.791 to -0.232 ($P$=.03); and b=-3.630, 95% CI -6.545 to -0.715 ($P$=.02). High education was also a significant determinant for a lower self-efficacy (b=-3.521, 95% CI -6.469 to -0.572; $P$=.02). Other determinants were not significant.

**Conclusions:** This study showed that the higher-educated users of a patient portal scored lower on usability and self-efficacy. Usability was also lower for older people and for patients with asthma or chronic obstructive pulmonary disease. The results portal is not tailored for different groups. Further research should investigate which factors from a patient's perspective are essential to tailor the portal for different groups and how a result portal can be optimally integrated within the daily practice of a doctor.

***(JMIR Form Res 2021;5(12):e25498)*** doi:10.2196/25498



XSL•FO
RenderX

**KEYWORDS**

patient portal; eHealth impact questionnaire; laboratory test results; self-efficacy; usability; age; gender; chronic disease; education; patient characteristics

## Introduction

The involvement of patients is important to allow them to manage their own health. When patients are more engaged, they tend to make better decisions on health behavior [1]. Patient involvement has increasingly been stimulated with digital possibilities [2], such as in patient portals [3,4]. A Dutch patient portal developed by Saltro Diagnostic Center provides patients access to laboratory test results, including explanatory information and visualization [3]. The aim of this portal is to increase patients' knowledge and to facilitate them to take an active role in their diagnostic process (eg, to ask questions and share opinions to improve the diagnostic process and reduce the risk of diagnostic errors [5]). Patient portals conveying laboratory test results in understandable language can help patients to take a more active role in managing their own health [6]. Therefore, it is recommended to test how patients perceive online portals and test results, for example by using the eHealth impact questionnaire (eHIQ) [7].

In 2019, we investigated patients' attitudes toward the same portal designed to communicate laboratory test results using the eHIQ [6]. The usability of this portal was rated positively, suggesting that the study participants found the patient portal easy to use, considered it trustworthy and appropriate, and that the provided information was easy to understand. The self-efficacy of the patients received a satisfying score, referring to whether patients were motivated to act on the presented information. It was concluded that the patients were generally positive toward the portal with opportunities to optimize self-efficacy; however, the impact of patient characteristics was not accounted for. Patient characteristics such as gender, age, education, and chronic disease can affect the usage of portals and the user satisfaction [8-10]. Limited research is available, specified for online communicating laboratory test results, on whether portal use and acceptance differ between groups. Further research on potential group differences is necessary to fine-tune the portal, making it acceptable for every user. We aim to replicate the previous study with larger numbers to examine how different groups of patients perceived the portal.

The main aim of this study is to evaluate the effect of gender, age, education, and chronic disease on the usability and self-efficacy of patients using a patient portal designed to communicate laboratory tests.

## Methods

### Design and Participants

A cross-sectional real-world study was conducted between December 2019 and July 2020 to explore the influence of patient characteristics on the usability of a patient portal and on self-efficacy. Patients who viewed their test results in the portal were automatically approached to complete the eHIQ. Age, gender, education level, and chronic disease were measured as well. There were no further inclusion or exclusion criteria.

No personal information was collected, and the data could not be traced back to the individual. Therefore, this study does not fall under the Medical Research Involving Human Subject Act (Wet medisch-wetenschappelijk onderzoek met mensen) and did not require approval from an ethics committee.

### Patient Portal

In 2015, Saltro launched a web-based portal that gives patients access to their own laboratory test results, including understandable explanatory information [3]. The content was created by a team of patients, general practitioners (GPs), communication specialists, and clinical chemists. Researchers estimated the level of health literacy of the information at communication level 1B based on the scales of the Common European Framework of Reference for Languages [11]. Daily, approximately 300 unique individuals look up their laboratory test results with the option to share their results with others.

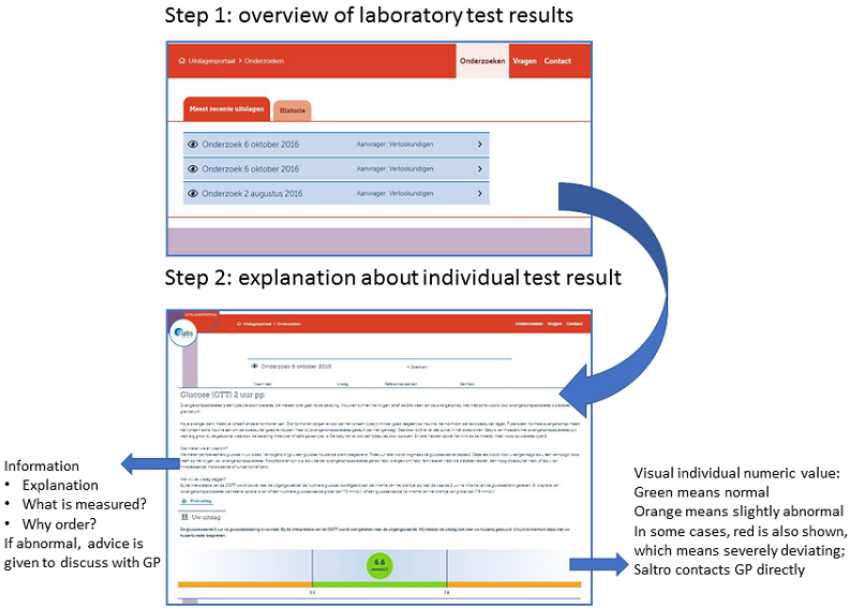
After blood withdrawal, the patients can look up their results by logging into the portal, with a username and password, through the website of the GP. The log-in procedure adheres to Dutch security legislation and guidelines (ie, the Dutch Personal Data Protection Act) and the General Data Protection Regulation guidelines. The patients can see an overview of all laboratory tests ordered by date (Figure 1). Each result has traffic-light–colored bullets and a visual.

This portal can be approached directly for laboratory test results but can also be approached within other portals as a plug-in; for example, a GP portal that functions as medication description.



XSL•FO
RenderX

**Figure 1.** Example of a test result with explanation. GP: general practitioner.




## Outcome Measures

Primary outcomes were "information and presentation" and "motivation and confidence to act" in the Dutch version of the eHIQ, part 2 (eHIQ2) [12,13]. The eHIQ2 is a self-reporting questionnaire measuring patients' attitudes toward a specific health-related website. Each of the 26 items is scored on a 5-point Likert scale ranging from "strongly disagree (1)" to "strongly agree (5)." The questionnaire has three subscales: information and presentation; motivation and confidence to act; and identification. The "information and presentation" subscale has 13 items and measures whether people find the website easy to use, which includes items on understanding, trustworthiness, and whether images used were appropriate. This subscale relates to usability. The "motivation and confidence to act" subscale consists of 10 items and assesses whether an individual felt reassured after reading the information on the website and was motivated to manage their health. This subscale relates to self-efficacy. The final subscale, identification, consists of 3 items and measures whether individuals identify with others who use the website. An example item is the following: "I feel I have a sense of solidarity with other people using the website." As users of the patient portal do not interact with other users, this subscale was considered irrelevant for the current study and is therefore not discussed further. The total scores per subscale were transformed to a 0-100 scale (higher scores representing a more positive attitude).

The determinants were age, gender, education, and chronic disease (Table 1), based on studies demonstrating that portal use was influenced inter alia by age, gender, presence of a chronic illness, education, and health literacy level [8,9,14]. Gender, instead of sex, was chosen because the patients' attitude and experience were analyzed. There was no biological measurement involved. Education level was chosen, but not health literacy, in order to minimalize the participants' number of questions. Relationships are proven between health literacy and education level, although health literacy is also common among the highly educated [15]. The choice for types of chronic diseases is based on the 5 most prevalent chronic diseases in the Netherlands: diabetes mellitus, asthma, chronic obstructive pulmonary disease (COPD), cardiovascular disease, and cancer [16,17]. Except for cancer, Saltro performs the blood test for these types of chronic diseases. People with asthma and COPD receive the same pulmonary function test and are therefore considered as 1 patient group in this research. Diabetes mellitus, Asthma or COPD, and cardiovascular diseases are the most prevalent chronic diseases in the population of Dutch GPs; these chronically ill patients are regularly monitored by GPs in a chronic care program with regular laboratory checks.



XSL•FO
RenderX

**Table 1.** Patient characteristics.

| Determinant | Variables |
|---|---|
| Age | Age at completing the questionnaire |
| **Gender** | |
| | Male |
| | Female |
| **Education** [18] | |
| | Low (no education, high school) |
| | Intermediate (intermediate vocational education) |
| | High (bachelor's degree, master's degree, doctorate) |
| **Chronic disease** | |
| | Diabetes mellitus |
| | Asthma or chronic obstructive pulmonary disease |
| | Cardiovascular disease |
| | None |

### Statistical Analyses

Descriptive analyses were performed to explore patient characteristics, usability, and self-efficacy. Univariable regression analyses were performed with age, gender, education, and chronic disease as determinants and usability and self-efficacy as outcomes. Significant ($P$<.10) determinants were included in multivariable models to examine which characteristics were independently related to the outcomes. To be rather inclusive than exclusive regarding the selection of variables for our multivariable model, $P$=.10 was chosen. A common level of $P$=.05 might fail to include relevant variables in those models [19]. For all other analyses and conclusion, $P$<.05 was considered statistically significant. The analyses were performed using the SPSS version 24 (IBM Corp) [20].

## Results

### Participant Characteristics, Usability, and Self-efficacy

The questionnaire was completed by 748 respondents. Response rate was 1.9% (39,430 unique visitors during the study period). The participants had a mean age of 58.5 years (SD 16.4), and they were mostly female (428/748, 57.2%) and highly educated (423/748, 56.6%) (Table 2). Moreover, 509/748 (68%) had no chronic disease. The mean scores of usability and self-efficacy were 68.9 (SD 10.6) and 62.5 (SD 13.1), respectively (Table 3). The mean (SD) scores on all items of the "information and presentation" and "motivation and confidence to act" domains can be found in Multimedia Appendix 1.



XSL•FO
RenderX

**Table 2.** Patient characteristics (N=748).

| Characteristics | Values |
|---|---|
| Age (years), mean (SD) | 52.8 (16.4) |
| **Gender, n (%)** | |
| Male | 314 (42.0) |
| Female | 428 (57.2) |
| Missing value | 6 (0.8) |
| **Education, n (%)** | |
| Low (no education, high school) | 96 (12.8) |
| Intermediate (intermediate vocational education) | 220 (29.4) |
| High (bachelor's degree, master's degree, doctorate) | 423 (56.6) |
| Missing value | 9 (1.2) |
| **Chronic disease, n (%)** | |
| Diabetes mellitus | 93 (12.4) |
| Asthma or COPD[a] | 54 (7.2) |
| Cardiovascular disease | 87 (11.6) |
| No chronic disease | 509 (68.0) |
| Missing value | 5 (0.7) |

[a]COPD: chronic obstructive pulmonary disease.

**Table 3.** Mean scores on the eHealth impact questionnaire (eHIQ); N=747.

| Subscale | Value |
|---|---|
| Usability[a], mean (SD) (1 missing[b]) | 68.9 (10.6) |
| Self-efficacy[c], mean (SD) (1 missing) | 62.5 (13.6) |

[a]Usability is measured with the eHIQ subscale "information and presentation".

[b]Missing value: one respondent gave the same answer to every question, including reversed questions, which indicates false responding.

[c]Self-efficacy is measured with the eHIQ2 subscale "motivation and confidence to act."

## Determinants for Perceived Usability

Age, education level, and chronic disease were relevant determinants with $P<.10$ for usability in the univariable analysis; they and where subsequently added in the multivariable model (Table 4). Multivariable analysis showed that higher age and high education were associated with a decreased usability: respectively, $b$=-.094, 95% CI -1147 to -0.042 ($P<.001$); and $b$=-2.512, 95% CI -4.791 to -0.232 ($P$=.03). Chronic disease affected usability, with patients with asthma or COPD scoring significantly lower compared with those without a chronic disease ($b$=-3.630, 95% CI -6.545 to -0.715; $P$=.02).



XSL•FO
RenderX

**Table 4.** Determinants for perceived usability.

| Determinant | | Univariable analysis | | Multivariable analysis | |
|---|---|---|---|---|---|
| Reference group | Determinant | *b*[a] (95% CI) | *P* value | *b* (95% CI) | *P* value |
| Age per year | Age | -.067 (-0.114 to -0.021) | .004 | -.094 (-1.147 to -0.042) | <.001 |
| Male | Gender | 1.322 (-0.234 to 2.878) | .10 | -.153 (-1.806 to 1.500) | .86 |
| Low education | Intermediate education | 1.275 (-1.224 to 3.774) | .32 | 1.262 (-1.189 to 3.712) | .31 |
| | High education | -1.992 (-4.302 to 0.318) | .09 | -2.512 (-4.791 to -0.232) | .03 |
| No chronic disease | Diabetes | -.377 (-2.692 to 1.939) | .80 | .347 (-2.053 to 2.747) | .78 |
| | Asthma or COPD[b] | -3.399 (-6.337 to -.416) | .02 | -3.630 (-6.545 to -0.715) | .02 |
| | Cardiovascular disease | -.286 (-2.668 to 2.096) | .81 | .890 (-1.576 to 3.357) | .48 |

[a]*b*: unstandardized beta value.

[b]CODP: chronic obstructive pulmonary disease.

### Determinants for Perceived Self-efficacy

Education level was a relevant determinant for self-efficacy in the univariable analysis with *P*<.10 (*b*=-3.521, 95% CI -6.469 to -.572; *P*=.02) (Table 5). Other determinants were not relevant; therefore, there was no need for a multivariable analysis.

**Table 5.** Determinants to perceived self-efficacy.

| Determinant | | Univariable analysis | |
|---|---|---|---|
| Reference group | Determinant | *b*[a] (95% CI) | *P* value |
| Age per year | Age | -.035 (-.095 to 0.024) | .24 |
| Male | Gender | -1.490 (-3.478 to 0.498) | .14 |
| Low education | Intermediate education | .159 (-3.031 to 3.348) | .92 |
| | High education | -3.521 (-6.469 to -0.572) | .02 |
| No chronic disease | Diabetes | -1.279 (-4.254 to 1.697) | .40 |
| | Asthma or COPD[b] | -2.438 (-6.214 to 1.338) | .21 |
| | Cardiovascular disease | 2.205 (-.856 to 5.265) | .16 |

[a]*b*: unstandardized beta value.

[b]CODP: chronic obstructive pulmonary disease.

## Discussion

### Principal Findings

This study aimed to evaluate the impact of patient characteristics on the perceived usability and self-efficacy of a patient portal. Higher education was associated with decreased usability and self-efficacy. Furthermore, usability was lower for older patients and for patients with asthma or COPD. The eHIQ is a validated questionnaire, and the results of this study with the eHIQ are in line with our previous study [6].

The finding that highly educated people have a significantly lower perceived usability and self-efficacy after using the portal is not in line with other research projects [21]. Mostly, people with high education tend to be more eHealth literate, showing more positive outcomes (motivation, self-efficacy, and better interaction with the doctor) after reading health information on the internet [21]. The use of qualitative interviews with the participants to explore the usability findings would be worthwhile. Nonetheless, other research projects on digital health information showed that tailoring—enabling users to self-tailor the preferred mode of information delivery via text and (audio)visuals—enhanced satisfaction with attractiveness and comprehensibility as compared with various versions of the nontailored digital information [22,23]. The patients were directly involved in the design phase of the studied results portal. However, the portal is not tailored for a specific group and might not be suitable for highly educated people. The continued development of the portal is an opportunity to take into account, especially by involving different education groups to give tailored advice through the portal.

This study also revealed that older participants scored lower on the usability of the portal. In other studies, the differences between age groups could be explained via the groups' digital skills. Van Deursen et al [24] and Broekhuizen et al [25] found that a higher age lowered operational and formal internet skills, such as operating an internet browser and maintaining a sense of orientation. However, in a study about the association of the usage of a public evidence-based health website and health care consultations, the use of digital information led to a decrease in regular doctors' consultations for older people in the same



XSL•FO
RenderX

way as for other age groups [26]. Nevertheless, the presentation and design of test results should be tailored for every age group [27] to obtain excellent usability and self-efficacy.

Furthermore, our research demonstrated that patients with asthma or COPD were more negative about usability. Other research projects reported that these patients are often insufficiently capable of understanding health information [28], which could be explained by anxiety, specific illness perception, age, and disease severity [29]. Other studies showed that the use of COPD self-management platforms is higher when the platform is an integrated part of health care [4]. Finally, some studies emphasized the importance of integrating skill-building activities into comprehensive education programs that enable patients with severe cases of asthma or COPD to identify high-quality sources of web-based health information [30]. Our study revealed that asthma or COPD patients are more negative about the results portal. Even more important for this group is tailoring the portal and integrating it into usual care [4]. Therefore, considerations for redesigning the online portal are at issue, together with COPD patients.

### Strength and Limitations

A strength of our study was the high sample size and that the patients completed the questionnaire immediately after they viewed their results, thereby limiting recall bias and giving an accurate picture of the patients' attitudes toward the portal. Nevertheless, those who completed the study questionnaire were a small portion of the total group that used the patient portal. The low response rate precludes generalizing whether the patient portal display and explanation of the results are acceptable and informative for all of the patients. In future research, it is interesting to compare patients that use the portal to those who do not. Moreover, we were not exhaustive with the possible patient characteristics as determinants. We cannot determine other factors that contribute to the patients' perceived usability and self-efficacy after seeing their lab results online. Possible other determinants that may impact usability and self-efficacy are the quality of the portal, the motivation to use the internet for health improvement [31], and the way patients use their knowledge in relation to the doctor [32,33]. Regarding the patient portal itself, lab results need to be easily understandable [34], and technology needs to be easy to use [8,35]. Previous research shows that the related lab results are easily understandable and that the patient portal is easy to use [3]. Therefore, it is interesting to explore which other factors influence a patient's attitude toward the patient portal.

### Conclusions

Highly educated users of a test results portal scored lower on usability and self-efficacy. The usability was also lower for older people and for patients with asthma or COPD. Result portals must adapt the language and communication used, according to the different target groups of age, education, and chronic illness. Only then can users take full advantage of the online information provision. Further research is necessary to determine promoting factors that users themselves consider important in a results portal, in order to tailor it for different groups. Further research is also needed on ways in which a portal can be optimally implemented and integrated within the daily practice of a doctor.

### Acknowledgments

The authors would like to thank the patients that were involved in several phases of the development of the portal and this study. The authors also thank the professionals and employees of Saltro who contributed to the development of the portal and this study.

### Conflicts of Interest

RT and AG are employees of Saltro, where the portal has been developed and implemented. ET is a former employee of Saltro.

Multimedia Appendix 1
Mean (SD) score subscales "information and presentation" and "motivation and confidence to act" from the eHealth impact questionnaire (eHIQ).
[DOCX File , 15 KB - formative_v5i12e25498_app1.docx ]

### References

1. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]
2. Snyder H, Engström J. The antecedents, forms and consequences of patient involvement: A narrative review of the literature. Int J Nurs Stud 2016 Jan;53:351-378. [doi: 10.1016/j.ijnurstu.2015.09.008] [Medline: 26602069]
3. Struikman B, Bol N, Goedhart A, van Weert JCM, Talboom-Kamp E, van Delft S, et al. Features of a Patient Portal for Blood Test Results and Patient Health Engagement: Web-Based Pre-Post Experiment. J Med Internet Res 2020 Jul 20;22(7):e15798 [FREE Full text] [doi: 10.2196/15798] [Medline: 32706704]
4. Talboom-Kamp EP, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJ, Numans ME, et al. High Level of Integration in Integrated Disease Management Leads to Higher Usage in the e-Vita Study: Self-Management of Chronic Obstructive Pulmonary Disease With Web-Based Platforms in a Parallel Cohort Design. J Med Internet Res 2017 May 31;19(5):e185 [FREE Full text] [doi: 10.2196/jmir.7037] [Medline: 28566268]



XSL•FO
RenderX

5. McDonald KM, Bryce CL, Graber ML. The patient is in: patient involvement strategies for diagnostic error mitigation. BMJ Qual Saf 2013 Oct;22 Suppl 2:ii33-ii39 [FREE Full text] [doi: 10.1136/bmjqs-2012-001623] [Medline: 23893394]
6. Talboom-Kamp E, Tossaint-Schoenmakers R, Goedhart A, Versluis A, Kasteleyn M. Patients' Attitudes Toward an Online Patient Portal for Communicating Laboratory Test Results: Real-World Study Using the eHealth Impact Questionnaire. JMIR Form Res 2020 Mar 4;4(3):e17060. [doi: 10.2196/17060]
7. Kelly L, Jenkinson C, Ziebland S. Measuring the effects of online health information for patients: item generation for an e-health impact questionnaire. Patient Educ Couns 2013 Dec;93(3):433-438 [FREE Full text] [doi: 10.1016/j.pec.2013.03.012] [Medline: 23598293]
8. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the Users of a Patient Portal: Cross-Sectional Study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
9. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
10. Rains SA. Perceptions of Traditional Information Sources and Use of the World Wide Web to Seek Health Information: Findings From the Health Information National Trends Survey. Journal of Health Communication 2007 Oct 11;12(7):667-680. [doi: 10.1080/10810730701619992] [Medline: 17934943]
11. Taalprofielen. www.SLO.nl. Amersfoort: SLO; 2021 Jun 02. URL: https://www.slo.nl/thema/vakspecifieke-thema/mvt/erk/ [accessed 2021-11-30]
12. Kelly L, Ziebland S, Jenkinson C. Measuring the effects of online health information: Scale validation for the e-Health Impact Questionnaire. Patient Educ Couns 2015 Nov;98(11):1418-1424 [FREE Full text] [doi: 10.1016/j.pec.2015.06.008] [Medline: 26162953]
13. Neijenhuijs KI, van der Hout A, Veldhuijzen E, Scholten-Peeters GGM, van Uden-Kraan CF, Cuijpers P, et al. Translation of the eHealth Impact Questionnaire for a Population of Dutch Electronic Health Users: Validation Study. J Med Internet Res 2019 Aug 26;21(8):e13408 [FREE Full text] [doi: 10.2196/13408] [Medline: 31452516]
14. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
15. van der Heide I, Wang J, Droomers M, Spreeuwenberg P, Rademakers J, Uiters E. The relationship between health, education, and health literacy: results from the Dutch Adult Literacy and Life Skills Survey. J Health Commun 2013;18 Suppl 1:172-184 [FREE Full text] [doi: 10.1080/10810730.2013.825668] [Medline: 24093354]
16. RIVM: Chronische ziekten. RIVM. Bilthoven: RIVM; 2020. URL: https://www.rivm.nl/rivm/kennis-en-kunde/expertisevelden/chronische-ziekten [accessed 2021-11-30]
17. Factsheet Chronische aandoeningen bij 40-plussers. Vektis. 2019 Jun 04. URL: https://www.vektis.nl/intelligence/publicaties/factsheet-chronische-aandoeningen-bij-40-plussers [accessed 2021-11-30]
18. Bevolking; hoogstbehaald onderwijsniveau en onderwijsrichting. CBS. 2011 Nov 16. URL: https://www.cbs.nl/nl-nl/cijfers/detail/82816ned?dl=3948A [accessed 2021-11-30]
19. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med 2008 Dec 16;3:17 [FREE Full text] [doi: 10.1186/1751-0473-3-17] [Medline: 19087314]
20. SPSS Statistics 24.0.0. IBM. 2016. URL: https://www.ibm.com/docs/en/spss-statistics/24.0.0 [accessed 2021-11-30]
21. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
22. Nguyen MH, Bol N, Lustria MLA. Perceived Active Control over Online Health Information: Underlying Mechanisms of Mode Tailoring Effects on Website Attitude and Information Recall. J Health Commun 2020 Apr 02;25(4):271-282. [doi: 10.1080/10810730.2020.1743797] [Medline: 32286923]
23. Nguyen MH, Smets EMA, Bol N, Loos EF, Van Weert JCM. How Tailoring the Mode of Information Presentation Influences Younger and Older Adults' Satisfaction with Health Websites. J Health Commun 2018;23(2):170-180. [doi: 10.1080/10810730.2017.1421729] [Medline: 29345531]
24. van Deursen AJ, van Dijk JA. Using the Internet: Skill related problems in users' online behavior. Interacting with Computers 2009 Dec;21(5-6):393-402. [doi: 10.1016/j.intcom.2009.06.005]
25. Broekhuizen T, Emrich O, Gijsenberg M, Broekhuis M, Donkers B, Sloot L. Digital platform openness: Drivers, dimensions and outcomes. Journal of Business Research 2021 Jan;122:902-914. [doi: 10.1016/j.jbusres.2019.07.001]
26. Spoelman WA, Bonten TN, de Waal MWM, Drenthen T, Smeele IJM, Nielen MMJ, et al. Effect of an evidence-based website on healthcare usage: an interrupted time-series study. BMJ Open 2016 Nov 09;6(11):e013166 [FREE Full text] [doi: 10.1136/bmjopen-2016-013166] [Medline: 28186945]
27. Tao D, Yuan J, Qu X. Presenting self-monitoring test results for consumers: the effects of graphical formats and age. J Am Med Inform Assoc 2018 Aug 01;25(8):1036-1046 [FREE Full text] [doi: 10.1093/jamia/ocy046] [Medline: 29762686]
28. Heins MJ, Spreeuwenberg P, Heijmans M. Leven met een longziekte in Nederland: cijfers en trends over de zorg- en leefsituatie van mensen met een longziekte. Nivel. 2016. URL: https://www.nivel.nl/nl/publicatie/leven-met-een-longziekte-nederland-cijfers-en-trends-over-de-zorg-en-leefsituatie-van [accessed 2021-11-30]



XSL•FO
RenderX

29. Korpershoek Y, Bos-Touwen ID, de Man-van Ginkel JM, Lammers J, Schuurmans MJ, Trappenburg J. Determinants of activation for self-management in patients with COPD. Int J Chron Obstruct Pulmon Dis 2016;11:1757-1766 [FREE Full text] [doi: 10.2147/COPD.S109016] [Medline: 27536087]
30. Stellefson ML, Shuster JJ, Chaney BH, Paige SR, Alber JM, Chaney JD, et al. Web-based Health Information Seeking and eHealth Literacy among Patients Living with Chronic Obstructive Pulmonary Disease (COPD). Health Commun 2018 Dec;33(12):1410-1424 [FREE Full text] [doi: 10.1080/10410236.2017.1353868] [Medline: 28872905]
31. Otokiti A, Williams KS, Warsame L. Impact of Digital Divide on the Adoption of Online Patient Portals for Self-Motivated Patients. Healthc Inform Res 2020 Jul;26(3):220-228 [FREE Full text] [doi: 10.4258/hir.2020.26.3.220] [Medline: 32819040]
32. Wildevuur SE, Simonse LW, Groenewegen P, Klink A. Information and communication technology enabling partnership in person-centred diabetes management: building a theoretical framework from an inductive case study in The Netherlands. BMJ Open 2019 Jun 16;9(6):e025930 [FREE Full text] [doi: 10.1136/bmjopen-2018-025930] [Medline: 31209085]
33. Otte-Trojel T, Rundall TG, de Bont A, van de Klundert J, Reed ME. The organizational dynamics enabling patient portal impacts upon organizational performance and patient health: a qualitative study of Kaiser Permanente. BMC Health Serv Res 2015 Dec 16;15:559 [FREE Full text] [doi: 10.1186/s12913-015-1208-2] [Medline: 26674529]
34. Garrido T, Jamieson L, Zhou Y, Wiesenthal A, Liang L. Effect of electronic health records in ambulatory care: retrospective, serial, cross sectional study. BMJ 2005 Mar 12;330(7491):581 [FREE Full text] [doi: 10.1136/bmj.330.7491.581] [Medline: 15760999]
35. Ros WJ. MyUMC Patient Portal: patient and professional perspectives. Int J Integr Care 2012 Jun 13;12(4):None. [doi: 10.5334/ijic.967]

## Abbreviations

**COPD:** chronic obstructive pulmonary disease
**eHIQ:** eHealth impact questionnaire
**GP:** general practitioner

*Edited by G Eysenbach; submitted 04.11.20; peer-reviewed by J van Delden, F Lau, T Risling; comments to author 27.11.20; revised version received 20.01.21; accepted 27.09.21; published 17.12.21.*

*Please cite as:*
*Tossaint-Schoenmakers R, Kasteleyn M, Goedhart A, Versluis A, Talboom-Kamp E*
*The Impact of Patient Characteristics on Their Attitudes Toward an Online Patient Portal for Communicating Laboratory Test Results: Real-World Study*
*JMIR Form Res 2021;5(12):e25498*
*URL: https://formative.jmir.org/2021/12/e25498*
*doi:10.2196/25498*
*PMID:34927593*

©Rosian Tossaint-Schoenmakers, Marise Kasteleyn, Annelijn Goedhart, Anke Versluis, Esther Talboom-Kamp. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# An Early Warning Mobile Health Screening and Risk Scoring App for Preventing In-Hospital Transmission of COVID-19 by Health Care Workers: Development and Feasibility Study

Ronald Mbiine[1], MD; Cephas Nakanwagi[2], BNUR; Herve Monka Lekuya[1], MD; Joan Aine[1], BSc; Kawesi Hakim[1], BSc; Lilian Nabunya[1], BSc; Henry Tomusange[1], BSc

[1]Makerere University College of Health Sciences, Kampala, Uganda

[2]Mulago National Referral Hospital, Kampala, Uganda

**Corresponding Author:**
Ronald Mbiine, MD
Makerere University College of Health Sciences
Mulago Hill Road
Kampala, 7072
Uganda
Phone: 256 774338585
Email: mbiineron@gmail.com

## Abstract

**Background:** Hospitals have been identified as very high-risk places for COVID-19 transmission between health care workers and patients who do not have COVID-19. Health care workers are the most at-risk population to contract and transmit the infection, especially to already vulnerable patients who do not have COVID-19. In low-income countries, routine testing is not feasible due to the high cost of testing; therefore, presenting the risk of uncontrolled transmission within non–COVID-19 treatment wards. This challenge necessitated the development of an affordable intermediary screening tool that would enable early identification of potentially infected health care workers and for early real time DNA–polymerase chain reaction testing prioritization. This would limit the contact time of potentially infected health care workers with the patients but also enable efficient use of the limited testing kits.

**Objective:** The aims of this study are to describe an early warning in-hospital mobile risk analysis app for screening COVID-19 and to determine the feasibility and user-friendliness of the app among health care workers.

**Methods:** The primary result of this research project was the development of a mobile-based daily early warning system for in-hospital transmission of COVID-19. Overall, the Early Warning System for In-Hospital Transmission of COVID-19 (EWAS) mobile app was found to be feasible, with over 69% of the health care workers having logged more than 67% of the required times. Over 93% of the participants reported that the tool was easy to use.

**Results:** The primary result of this research project was the development of a mobile-based daily early warning system for in-hospital transmission of COVID-19. Overall, the Early Warning System for In-Hospital Transmission of COVID-19 (EWAS) mobile app was found to be feasible, with 69% of the health care workers (69/100) having logged more than 67% of the required times. Of the 100 participants, 93 reported that the tool was easy to use.

**Conclusions:** The EWAS mobile app is a feasible and user-friendly daily risk scoring tool for preventing in-hospital transmission of COVID-19. Although it was not designed to be a diagnostic tool but rather a screening tool, there is a need to evaluate its sensitivity in predicting persons likely to have contracted COVID-19.

***(JMIR Form Res 2021;5(12):e27521)*** doi:10.2196/27521

**KEYWORDS**

mHealth; risk score for Covid-19; Africa; mobile health; digital health; pandemic; COVID-19; COVID; screening tool; healthcare workers; transmission; warning system



XSL•FO
RenderX

## Introduction

COVID-19 continues to present a serious national and global challenge [1]. In Uganda, the disease is now at the community transmission stage, implying that all mitigation measures such as contact tracing are no longer feasible.

One of the key identified areas of superspread is the hospitals where patients who do not have COVID-19 go to seek care [2,3]. Health care workers, by the nature of their work, are at the highest risk of contracting and transmitting COVID-19 [4]. Its estimated that over 10,000 health care workers in Africa have been infected with COVID-19 [5]; meanwhile, a significant number of vulnerable patients contract COVID-19 while admitted to hospitals for other ailments [6]. To minimize the in-hospital transmission of COVID-19, especially by health care providers who interact with several patients, it would be necessary to conduct routine DNA–polymerase chain reaction (PCR) tests for COVID-19 among these health care providers. The cost of a DNA-PCR COVID-19 test, which is approximately US $50 in Uganda [7], coupled with the 24- to 48-hour turnaround time for the results makes this test an unsustainable screening method for early prevention of in-hospital disease transmission in the medium term, especially in low-income country hospital settings.

This therefore presents a great challenge in early detection of COVID-19, creating the need for an intermediary screening tool that can easily identify potentially sick health care workers based on a daily log of their symptoms and contact history.

The Early Warning System for In-Hospital Transmission of COVID-19 (EWAS) is a mobile health (mHealth)–based app that is designed purposefully to address this gap. This application enables the daily screening and early identification of potentially infected health care workers for prioritization in testing; this will enable better use of the limited DNA-PCR test kits, as priority will be given to testing those with a high risk score. This reduces the need to conduct regular mass DNA-PCR testing of all staff, which is neither feasible nor sustainable. The mobile app thus creates an intermediary risk-scoring tool that is self-administered and very simple, and it could quickly identify a potentially sick health care worker long before they suspect that they have COVID-19 or obtain their test result.

Owing to the ambiguous nature of COVID-19 symptoms, infected persons are usually initially skeptical about having contracted COVID-19; it is only after they have developed the classic symptoms that they may proceed to self-isolate and undergo a test.

Often during this vague prodromal period, the virus is already transmissible, and the unsuspecting health care worker may still be delivering care. With a daily screening mobile app, the presence of these vague symptoms may be identified, enabling the instillation of the necessary mitigation measures on the ward.

The mobile app also has the ability to predict cluster epidemics long before they would be observed, as a collective high score among multiple health care workers on one of the wards would imply this possibility.

This tool therefore complements DNA-PCR testing as an adjunctive intermediary screening tool and does not replace or intend to replace the existing testing as recommended by the Ministry of Health of Uganda. Therefore, the intention of the study is not to see how sensitive the app is in diagnosing COVID-19; as this was never the purpose of its development, there is no intention to evaluate its sensitivity or specificity.

The primary research question for this study was to determine the feasibility and acceptance of using the EWAS mobile app in reducing hospital transmission of COVID-19.

We hypothesize that the EWAS app is a feasible and acceptable intermediary screening tool for preventing transmission of COVID-19 on non–COVID-19 treatment wards.

The objectives of this study where therefore to create an intermediary daily early warning COVID-19 risk scoring screening mobile tool and evaluate its feasibility and acceptability of use on the clinical wards in Mulago Hospital.

The intended aim would be to prioritize health care workers with high risk scores for DNA-PCR testing and timely intervention. In this way, the exposure time of potentially infected health care workers to patients would be significantly reduced, thus minimizing the in-hospital transmission of COVID-19 on general and non–COVID-19 treatment units.

In this study, first, we describe the mobile app in detail, and we subsequently describe the feasibility and acceptability findings for its use.

## Methods

This study constituted of two phases. The first phase was the design and development of the software app, and the second phase included the prospective piloting of the mobile app.

### Study Setting

The study was conducted in the Directorate of Surgical Services of the Mulago National Referral Hospital. This is the largest national referral hospital in Uganda, based in Kampala.

The Directorate of Surgery comprises five surgical departments: neurosurgery, gastrointestinal surgery, cardiothoracic surgery, accident and emergency, and breast and endocrine.

These wards treat surgical conditions and are not designated COVID-19 treatment units; the average monthly patient turnover is 300.

### Study Population

The study participants included health care workers who directly participated in patient care in the different departments in the Directorate of Surgical Services. These include surgeons, surgical residents, nurses, and intern physicians.

### Sampling and Sample Size Estimation

This aim of this study was to assess the feasibility of using this app, and it is the first of its kind.

Because it was a pilot study, the sample size was based on the convenience of a workable number for the stipulated study period of 1 month. Therefore, 20 participants were randomly



XSL•FO
RenderX

selected from each of the five departments in the Directorate of Surgery to constitute a sample size of 100 participants.

To be eligible for inclusion in the study, participants were required to have a smart mobile phone with internet connectivity. Because the first version of the app was built only for Android phones, participants with iPhones were excluded from the study.

## Outcome Variables

The primary outcome variables for the study were as follows:

1. Feasibility of regular use: For this study, feasibility of regular use was determined as the mean number of days on which users logged their risk scores into the mobile app. Logging at least 20 out of the 30 daily scores was ranked as "feasible for regular use."
2. User-friendliness: This was defined based on the participants' rating of how easy it was to understand and use the mobile app. This was based on a 5-point Linkert scale, with the highest score of 5 for "very user-friendly" and zero for "not user-friendly."

## Study Procedure

Ethical approval and administrative clearance were obtained from the Mulago National Referral Hospital ethics review board. Upon clearance, the project was implemented in 2 phases. The first phase was the app development, and the second phase was the feasibility assessment.

## Application Development

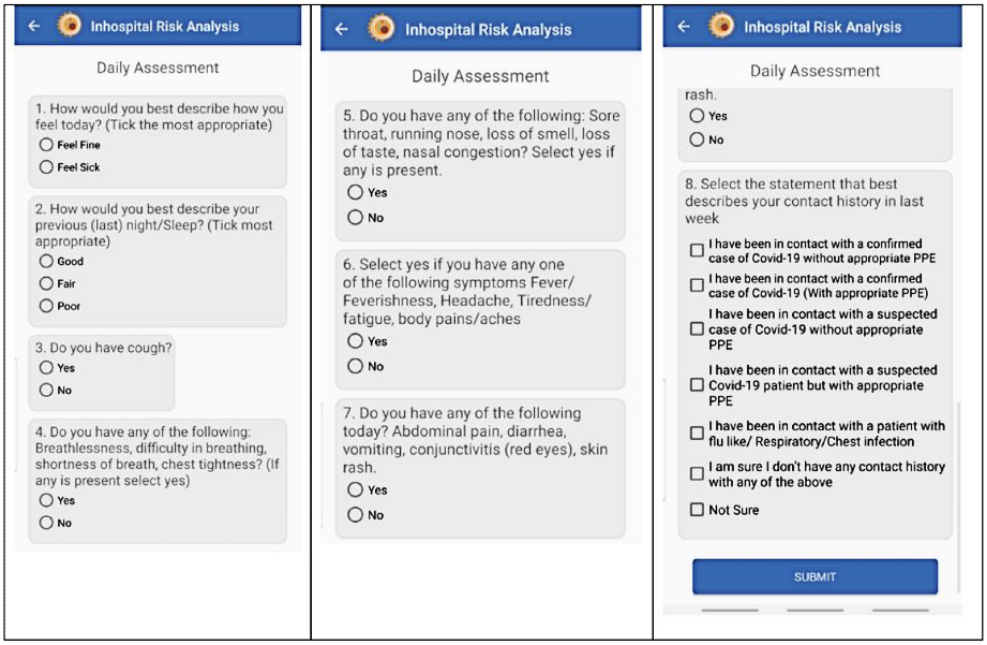
The development of the EWAS mobile app comprised two subphases: (1) risk assessment tool and algorithm development and (2) software design and development.

## Risk Screening Tool and Algorithm Development

To develop a standard risk assessment tool, we identified the World Health Organization (WHO) [8] and Ministry of Health of Uganda [9] guidelines for the clinical case definition of COVID-19.

Based on the above case definitions, we developed an 8-question risk assessment and scoring tool comprising multiple-choice answers, of which the participant could only select a single answer for each question (see Figure 1). This was done to minimize user fatigue from the long examination-style assessment tools that are commonly used for screening.

**Figure 1.** Screenshots of the risk assessment tool in the Early Warning System for In-Hospital Transmission of COVID-19 (EWAS) app.




Each selected answer was assigned a score based on whether it comprised a positive symptom for COVID-19.

The 8 questions were grouped into 2 sections; section I comprised the symptom score, which was based on 7 questions evaluating the user's general health condition as well as the presence of COVID-19–like symptoms. The overall maximum score for section I was 155, with the minimum score being 30.

Section II comprised the contact score, consisting of 1 multiple-choice question describing possible scenarios for the contact history of the user in the last week. Each answer was weighted, with the maximum score being 50 and the minimum being 15.

To include potential contact with asymptomatic carriers, even when the participant was sure that they had not been in contact with a suspected COVID-19–positive patient, they could not be awarded a score of 0.

To avoid falsely alarming high risk scores, each of the 8 questions was weighted based on how prevalent the given symptoms were among patients with COVID-19 in Uganda [10].

Section I and section II each contributed 50% to the overall risk score, which was calculated as the percentage of the participant score relative to the total score.



XSL•FO
RenderX

Hence, to calculate the overall patient risk score, the patient symptom score and the contact score would be entered into the following formula:

Patient overall risk score = [(symptom score ÷ 155) × 50] + [(contact score ÷ 50) × 50]

Based on the daily input, a daily risk score would be awarded to the participant (Figure 2, left).

The second step was to calculate a trend analysis of the daily logged risk score.

Based on the weekly trends of the risk scores displayed in a graphical pattern (Figure 2, center), the participant was issued a weekly risk badge (Figure 2, right). The risk badge would generally reflect the risk status of the patient regarding their COVID-19 symptoms and contact score. The intention of the risk badge was to identify participants who had a high-risk badge and prioritize them for a weekly DNA-PCR test.

Hence, if a participant consistently had a high risk score through the week, their risk badge would identify them as high risk and indicate the necessity for urgent DNA-PCR testing.

The app also identified clusters of participants with high-risk badges, and if these were on one ward, this would quickly highlight the affected ward as having potential for a cluster outbreak within the hospital; this feature was implemented with the intention of enabling timely intervention before the lives of patients and other health care workers were placed at high risk.

Once the above risk assessment tools and algorithms were completed, they were reviewed by a clinic epidemiologist and a physician who was directly participating in the treatment of patients with COVID-19 at Mulago National Referral Hospital, and after they cleared the review process, ethical clearance was obtained from the Research and Ethics Review Board of Mulago Hospital (protocol number MHREC 1901).

Subsequently, the software developer team undertook the development of the mobile app with multiple pretests to ensure good end-user experience and relevancy.

To improve user-friendliness, the app's risk assessment tool was designed to be as simple as possible, with minimal dropdown menus and no direct user input.

Because English is the official language of communication between health care workers while delivering care, the app was developed to be used only in English, and all the health care workers were sufficiently proficient in speaking, reading, and writing in the English language.

The app was piloted for 1 week among 10 participants, and the user feedback was used to make final adjustments prior to the official launch [11].

The app was later published on the Google Play store as "In-hospital Early Risk Analysis for COVID-19" [12].

**Figure 2.** Screenshots showing the daily risk score and weekly risk badge in the Early Warning System for In-Hospital Transmission of COVID-19 (EWAS) app.

## Availability and Requirements

The availability and requirements of the app are listed in Textbox 1.



XSL•FO
RenderX

**Textbox 1.** The availability and requirements of the Early Warning System for In-Hospital Transmission of COVID-19 (EWAS) app.

**Project name**: Early Warning System for In-Hospital Transmission of COVID-19 (EWAS)

**Project home page**: https://ewasinhospital.org/ [11]

**Google Play address**: https://play.google.com/store/apps/details?id=org.ewasinhospital.app.mobile [12]

**Operating system**: Android, platform independent

**Programming languages**: Java, Python, Django

**Other requirements**: Android 5+, Python 3, MySQL

**License**: none

**Any restrictions to use by nonacademics**: none

### Study Implementation

Following the selection of the 100 participants, user training was provided and the mobile app was installed on their devices. The participants were required to register in the software app prior to use, and their profession (position) and ward were captured. The participants were instructed to use the software on a daily basis for a period of 30 days. Participants were to log their daily risk scores and weekly risk badges and to share them with the research assistants. In the event that they had a high-risk badge, they were advised to contact COVID-19 surveillance for further assessment.

At the end of the study duration, the participants were required to answer a questionnaire to evaluate their user experience. The participants were also invited to provide information on how the app could be made more relevant in subsequent versions.

## Results

### Participant Demographics

In total, 100 participants were enrolled in the study, comprising 26 (26%) junior house officers (intern physicians), 28 (28%) nurses, 24 (24%) senior house officers (surgical residents), and 22 (22%) surgeons.

### Distribution of Staff in the Different Surgical Unit

The participants were distributed in different units, as shown in Table 1, with the majority being in the gastrointestinal surgical unit.

**Table 1.** Ward distribution of the health care workers who participated in the study (N=100).

| Unit | Participants, n (%) |
|---|---|
| Accident and emergency | 20 (20) |
| Breast and endocrine | 18 (18) |
| Cardiothoracic surgery | 16 (16) |
| Gastrointestinal surgery | 26 (26) |
| Neurosurgery | 20 (20) |

### Participant Mean Risk Scores Per Week

In the first week, 24% of the participants (24/100) had a high risk score, while in the second, third, and fourth weeks, 26, 19, and 23 of the 100 participants reported a high risk score, respectively (Table 2).

Table 3 shows the proportions of participants with high risk scores, as in, participants with risk scores greater than 75% in each week.

**Table 2.** Weekly mean risk scores of the participants.

| Week | Risk score, mean (SD) |
|---|---|
| 1 | 51.5 (14) |
| 2 | 53 (14.68) |
| 3 | 54.27 (14.06) |
| 4 | 50.32 (13.92) |



XSL•FO
RenderX

**Table 3.** Participants with high weekly risk scores per week (N=100).

| Week | High risk scores, n (%) |
|---|---|
| 1 | 11 (11) |
| 2 | 12 (12) |
| 3 | 14 (14) |
| 4 | 10 (10) |

### Feasibility of Regular Use/Compliance

The mean number of days that were logged over the 30-day period was 21.97 (SD 5.99); 69% of the participants (69/100) logged their data more than 20 times, which demonstrates high compliance and feasibility of use of the mobile app.

### Comparison of Regular Use Between Nurses, Junior House Officers, Senior House Officers, and Surgeons

Overall, the intern physicians demonstrated more use and compliance, with 73% (19/26) having logged and registered their scores on at least 20 of the 30 days; this was followed by the nurses, with 71% compliance (20/28). The surgeons had the lowest compliance among all the participants (Table 4).

**Table 4.** Feasibility of regular app use by profession.

| Professional background | Compliance, n (%) | |
|---|---|---|
| | Low | High |
| Junior house officers (n=26) | 7 (27) | 19 (73) |
| Senior house officers (n=24) | 8 (33) | 16 (67) |
| Nurses (n=28) | 8 (29) | 20 (71) |
| Surgeons (n=22) | 8 (36) | 14 (64) |
| Total (N=100) | 31 (31) | 69 (69) |

### User Experience

Of the 100 participants, 93 described the app as very easy to use, while 3 described the app as easy to use and 4 described the application as not easy, difficult, or very difficult to use. User errors or app crashing was reported by only 2 of the 100 participants, with 98 reporting no history of app crashes or errors.

The COVID-19 DNA-PCR testing rates were very low in the month under review, with only 7% (7/100) of the staff having undertaken a COVID-19 test.

Of the 100 participants, 74 (74%) were willing to continue using the app even after the end of the study.

### Efficacy in Predicting Positive COVID-19 Infection

At the time of the study implementation, the project had not been incorporated into the Ministry of Health protocols for managing COVID-19; therefore, these protocols could not independently inform the need to request a COVID-19 PCR test. Second, the unavailability of the DNA-PCR test for regular disease screening, coupled with the associated cost of $50 USD for each test, resulted in very low testing rates among the health care providers, with less than 10 having undertaken the DNA-PCR test during the period under review. The study could not therefore compare the efficacy of the tool in predicting potentially high-risk participants.

The primary intended outcome of this research project was to create an intermediary early warning tool that would help identify symptomatic health care workers to prevent in-hospital transmission of COVID-19 while also identifying health care workers who would require a DNA-PCR test without necessarily having to routinely perform mass testing of all health care workers, which is costly and unsustainable.

With the mobile app in place, health care workers are now able to obtain a daily risk score that is objective and quantifiable. This daily symptom score, when used, can help identify the disease process long before the health care worker would think that they had COVID-19, hence limiting the contact time of a potentially infected health care worker with a patient.

The health care worker can also access a graphical representation of their risk trends for the past week and month. This risk trend is the basis of the risk badge (Figure 2) that is assigned to the health care worker every after 5 days. The risk badge therefore forms the basis for the need for further analysis and intervention including DNA-PCR testing and mitigation measures, such as self-isolation.

The software also enables notification and messaging, through which an affected health care worker can send out an urgent request for support from the COVID-19 treatment team, including a request for a dispatch team. The app also comprises a repository for maintained and tracking COVID-19 DNA-PCR test results, which the user uploads onto their database.

## Discussion

### Principal Findings

In-hospital transmission of COVID-19 among health care workers and patients is still a recognized challenge in prevention



XSL•FO
RenderX

of COVID-19 transmission [13], especially to already vulnerable populations. The most ideal way of minimizing transmission would be routinely performing point-of-care DNA-PCR testing of all health care workers and patients on various non–COVID-19 treatment units, such as surgical wards [14]. In the study period, the user participants had a <7% (7/100) testing rate for COVID-19. This further illustrates the unsustainability of DNA-PCR testing as a regular screening tool for preventing COVID-19 infection [15,16]. In our study setting, the prohibitive US $50 out-of-pocket cost reduced the likelihood of regular testing as an alternative for preventing in-hospital transmission of COVID-19 [7].

Similarly, regular routine DNA-PCR testing is globally recognized as unfeasible, especially in low- and middle-income countries; therefore, in-hospital transmission of COVID-19 remains a substantial challenge [17,18].

Therefore, there was a need for an easily accessible self-administered daily risk scoring tool, which led to the development of this mHealth solution. One of the important concerns related to mHealth apps is whether they are feasible and can consistently be used unsupervised on a daily basis.

Overall, we found that the EWAS mobile app was feasible to use, based on the fact that the consistent use by 69% of the participants was similar to findings in other settings where digital screening tools were adapted for early detection of COVID-19 [19,20].

Interestingly, we found that intern physicians and nurses were more likely to consistently use the app when compared to surgical residents and surgeons. Unsurprisingly, digital utilization is more commonly adapted among younger medical professionals [21]; this explains the comparably high use among junior house officers, including junior physicians. Similarly, nurses have been described worldwide as embracing and using digital health applications [22,23]; this finding is consistent with the increased uptake and use of the mobile app by nurses in our study population. Intern physicians and nurses spend more contact time on the wards and are hence more likely to contract COVID-19 infection. Their increased ward contact in comparison to that of surgeons could also explain why the surgeons' consistency of use was the lowest.

Our study demonstrated that a mobile app for daily risk monitoring of COVID-19 is feasible and could consistently be used use a daily risk scoring tool for preventing in-hospital transmission of COVID-19.

The EWAS app for in-hospital transmission of COVID-19 can therefore be implemented as a daily screening tool to reduce the transmission rates on hospital wards by identifying health care workers who are potentially infected with the disease and therefore may be transmitting it. The tool is not designed to replace or be a diagnostic tool but to screen health care workers who may be prioritized for testing.

At the time of developing and implementing this project, there was no established local screening tool that would assign a risk score to a participant based on their symptoms. Patients would be evaluated for COVID-19–defining symptoms, and if these were classic based on the WHO [8] and Uganda government [9] guidelines, a test would then be recommended. These evaluations were not intended to identify potentially infected persons but identify patients for treatment.

Therefore, there was no comparative gold standard risk assessment screening tool that our app could be compared to; hence, it was necessary to perform a pilot study.

In highly transmissible diseases such as COVID-19, mHealth apps have been described to significantly improve disease screening and symptom monitoring [24-26]. Daily screening tools have an important role to play in preventing COVID-19 [27,28], as they may identify an infected person before they actually become suspicious of being infected.

In China, mHealth apps were shown to significantly reduce disease transmission [29]; however, these apps were designed for use in community settings. In Sweden, a similar mHealth app that tracks logistic use, including personal protective equipment, and patient care in 5 hospitals has been developed and is in use [30]. Overall, use of mHealth apps in disease tracking and symptom monitoring as an adjunct to existing guidelines in the management of COVID-19 is increasing [31].

In Africa, there has equally been a growing interest in using digital health technologies as surveillance or treatment monitoring tools [32]; however, no studies evaluating the use of mHealth solutions in preventing in-hospital transmission have been described.

The availability of this mHealth self-administered screening and risk assessment software in low-income countries may be relevant in the prevention of day-to-day disease transmission. This tool represents a first-line in-hospital active screening tool that is capable of identifying potentially infected health care workers before they transmit the illness. Without this app in our health care system, in-hospital transmission is essentially left to the discretion of the affected health care worker, and this represents a potential loophole for limiting disease transmission.

It should be noted that the EWAS mobile app does not replace testing or any other established guidelines in the diagnosis of COVID-19; however, it can be used in combination as an intermediary adjunct with these existing guidelines to increase the likelihood of minimizing the in-hospital transmission of COVID-19.

By being able to identify the health care workers most likely to be infected, the EWAS app can therefore maximize the potential benefits in screening and prioritizing the limited testing resources of DNA-PCR for health care workers in large hospital settings, such as the Mulago National Referral Hospital.

Despite the advantages of the EWAS app, health care provider compliance remains a great challenge, especially as the pandemic progresses. As the general fear associated with COVID-19 has waned, so has the compliance with the majority of established standard operating procedures, including compliance with the EWAS app.

This complacency is partly responsible for the recent spikes in the spread of COVID-19; however the increasing spread calls



XSL•FO
RenderX

for more aggressive campaigns and adaption of all mitigating measures, including the adoption of the EWAS tool in non–COVID-19 treatment units.

Moreover, for all health care workers to successfully use these tools, they would need to have access to a smartphone as well as a mobile data connection. In Uganda, smartphone coverage is approximately 42% [33], with over 70% smartphone coverage in urban centers. In Mulago Hospital, based on a preliminary survey, smartphone coverage was approximately 90% among health care workers; however, despite this good coverage, the ability to fully use the smartphones, including installation of the app and registration of the user, required the establishment of support services. To a great extent, this may limit the realization of the full potential of integrating mobile apps, including the EWAS app, in disease surveillance and transmission prevention.

## Conclusion

The EWAS mobile app is feasible and can consistently be used as a daily risk-scoring tool to prevent in-hospital transmission of COVID-19 among health care workers.

## Limitations

One of the key limitations of this study is the inability to compare the efficacy of this tool in predicting COVID-19 positivity among the participants. This was due to the inaccessibility to routine testing among participants due to the prohibitive fee of US $50 associated with the DNA-PCR testing.

## Recommendations

We recommend that the tool be further evaluated on a larger scale and in different hospital settings to compare its acceptance and consistence of use. We also recommend that the sensitivity of this tool in predicting a positive infection of COVID-19 compared to DNA-PCR testing be determined. This will help evaluate how good a screening tool the app is and advise on revisions that can be made to greatly increase its sensitivity.

## Data Availability

Data sharing is not applicable to this article, as no data sets were generated or analyzed during the case report.

## Acknowledgments

We wish to acknowledge the Makerere University Research Innovation Fund (MakRIF) and the Government of Uganda for the financial support necessary to develop and implement this project. We also thank the health care providers, including doctors and nurses, in the Directorate of Surgical Services of Mulago National Referral Hospital. Funding support was obtained from MakRIF. MakRIF was not involved in the writing, implementation, analysis, or interpretation during the project.

## Authors' Contributions

RM wrote the manuscript. CN critically revised the manuscript. HML critically reviewed the manuscript. HK, TM, and NL directly participated in the design and building of the software app and also revised the manuscript. All authors were directly involved in the implementation of the project.

## Conflicts of Interest

None declared.

## References

1. Baloch S, Baloch MA, Zheng T, Pei X. The coronavirus disease 2019 (COVID-19) pandemic. Tohoku J Exp Med 2020 Apr;250(4):271-278 [FREE Full text] [doi: 10.1620/tjem.250.271] [Medline: 32321874]
2. Majra D, Benson J, Pitts J, Stebbing J. SARS-CoV-2 (COVID-19) superspreader events. J Infect 2021 Jan;82(1):36-40 [FREE Full text] [doi: 10.1016/j.jinf.2020.11.021] [Medline: 33245943]
3. Frieden TR, Lee CT. Identifying and interrupting superspreading events-implications for control of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 2020 Jun;26(6):1059-1066 [FREE Full text] [doi: 10.3201/eid2606.200495] [Medline: 32187007]
4. Pruc M, Golik D, Szarpak L, Adam I, Smereka J. COVID-19 in healthcare workers. Am J Emerg Med 2021 Jan;39:236 [FREE Full text] [doi: 10.1016/j.ajem.2020.05.017] [Medline: 32414525]
5. Over 10 000 health workers in Africa infected with COVID-19. World Health Organization. 2020 Jul 23. URL: https://www.afro.who.int/news/over-10-000-health-workers-africa-infected-covid-19 [accessed 2020-07-23]
6. Power J. More than 200 patients infected with Covid-19 in hospital in four weeks. The Irish Times. 2020 Dec 21. URL: https://www.irishtimes.com/news/health/more-than-200-patients-infected-with-covid-19-in-hospital-in-four-weeks-1.4443325 [accessed 2021-11-30]
7. Athumani H. Uganda cuts cost of COVID test From $65 to $50. VOA. 2020 Oct 24. URL: https://tinyurl.com/2p9ffs4k [accessed 2021-11-30]
8. WHO COVID-19 case definition. World Health Organization. 2020 Dec 16. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance_Case_Definition-2020.2 [accessed 2021-11-30]



XSL•FO
RenderX

9. National guidelines for management of COVID-19. Ministry of Health, Republic of Uganda. URL: https://www.health.go.ug/cause/national-guidelines-for-management-of-covid-19/ [accessed 2021-11-30]
10. Kirenga B, Muttamba W, Kayongo A, Nsereko C, Siddharthan T, Lusiba J, et al. Characteristics and outcomes of admitted patients infected with SARS-CoV-2 in Uganda. BMJ Open Respir Res 2020 Sep 07;7(1):e000646 [FREE Full text] [doi: 10.1136/bmjresp-2020-000646] [Medline: 32900781]
11. EWAS: In-hospital Early Risk Analysis for Covid-19. URL: https://ewasinhospital.org/
12. In-hospital Early Risk Analysis for COVID-19. Google Play Store. URL: https://play.google.com/store/apps/details?id=org.ewasinhospital.app.mobile [accessed 2021-11-30]
13. Çelebi G, Pişkin N, Çelik Bekleviç A, Altunay Y, Salcı Keleş A, Tüz MA, et al. Specific risk factors for SARS-CoV-2 transmission among health care workers in a university hospital. Am J Infect Control 2020 Oct;48(10):1225-1230 [FREE Full text] [doi: 10.1016/j.ajic.2020.07.039] [Medline: 32771498]
14. Zhu H, Zhang H, Ni S, Korabečná M, Yobas L, Neuzil P. The vision of point-of-care PCR tests for the COVID-19 pandemic and beyond. Trends Analyt Chem 2020 Sep;130:115984 [FREE Full text] [doi: 10.1016/j.trac.2020.115984] [Medline: 32834243]
15. Treibel TA, Manisty C, Burton M, McKnight Á, Lambourne J, Augusto JB, et al. COVID-19: PCR screening of asymptomatic health-care workers at London hospital. Lancet 2020 May;395(10237):1608-1610. [doi: 10.1016/s0140-6736(20)31100-4]
16. Fennelly K, Whalen CC. Asymptomatic health-care worker screening during the COVID-19 pandemic. Lancet 2020 Oct 31;396(10260):1393 [FREE Full text] [doi: 10.1016/S0140-6736(20)32214-5] [Medline: 33129386]
17. Afzal A. Molecular diagnostic technologies for COVID-19: limitations and challenges. J Adv Res 2020 Nov;26:149-159 [FREE Full text] [doi: 10.1016/j.jare.2020.08.002] [Medline: 32837738]
18. Schultz MJ, Gebremariam TH, Park C, Pisani L, Sivakorn C, Taran S, COVID–LMIC Task Forcethe Mahidol–Oxford Research Unit (MORU), Bangkok, Thailand. Pragmatic recommendations for the use of diagnostic testing and prognostic models in hospitalized patients with severe COVID-19 in low- and middle-income countries. Am J Trop Med Hyg 2021 Jan 22 [FREE Full text] [doi: 10.4269/ajtmh.20-0730] [Medline: 33534752]
19. Ekong I, Chukwu E, Chukwu M. COVID-19 mobile positioning data contact tracing and patient privacy regulations: exploratory search of global response strategies and the use of digital tools in Nigeria. JMIR mHealth uHealth 2020 Apr 27;8(4):e19139 [FREE Full text] [doi: 10.2196/19139] [Medline: 32310817]
20. Shah SS, Gvozdanovic A, Knight M, Gagnon J. Mobile app-based remote patient monitoring in acute medical conditions: prospective feasibility study exploring digital health solutions on clinical workload during the COVID crisis. JMIR Form Res 2021 Jan 15;5(1):e23190 [FREE Full text] [doi: 10.2196/23190] [Medline: 33400675]
21. Payne KFB, Wharrad H, Watts K. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012 Oct 30;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]
22. Mayer MA, Rodríguez Blanco O, Torrejon A. Use of health apps by nurses for professional purposes: web-based survey study. JMIR mHealth uHealth 2019 Nov 01;7(11):e15195 [FREE Full text] [doi: 10.2196/15195] [Medline: 31682587]
23. Jedamzik S. Digital health and nursing: the future is now. Article in German. Unfallchirurg 2019 Sep;122(9):670-675. [doi: 10.1007/s00113-019-0672-2] [Medline: 31143981]
24. John Leon Singh H, Couch D, Yap K. Mobile health apps that help with COVID-19 management: scoping review. JMIR Nurs 2020;3(1):e20596 [FREE Full text] [doi: 10.2196/20596] [Medline: 32897271]
25. Alwashmi MF. The use of digital health in the detection and management of COVID-19. Int J Environ Res Public Health 2020 Apr 23;17(8) [FREE Full text] [doi: 10.3390/ijerph17082906] [Medline: 32340107]
26. R Niakan Kalhori S, Bahaadinbeigy K, Deldar K, Gholamzadeh M, Hajesmaeel-Gohari S, Ayyoubzadeh SM. Digital health solutions to control the COVID-19 pandemic in countries with high disease prevalence: literature review. J Med Internet Res 2021 Mar 10;23(3):e19473 [FREE Full text] [doi: 10.2196/19473] [Medline: 33600344]
27. Sung J, Choudry N, Bachour R. Development and validation of a simple risk score for diagnosing COVID-19 in the emergency room. Epidemiol Infect 2020 Nov 13;148:e273 [FREE Full text] [doi: 10.1017/S0950268820002769] [Medline: 33183384]
28. Shen SP, Wei YY, Zhao Y, Jiang Y, Guan JX, Chen F. Risk assessment of global COVID-19 imported cases into China. Article in Chinese. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Oct 10;41(10):1582-1587. [doi: 10.3760/cma.j.cn112338-20200415-00577] [Medline: 32455514]
29. Wu J, Xie X, Yang L, Xu X, Cai Y, Wang T, et al. Mobile health technology combats COVID-19 in China. J Infect 2021 Jan;82(1):159-198 [FREE Full text] [doi: 10.1016/j.jinf.2020.07.024] [Medline: 32730998]
30. Dress J. Swedish Health Services taps Microsoft to build app that tracks COVID-19 patients, hospital capacity. Becker's Hospital Review. Sweden: Becker's Health Care; 2020 Apr 02. URL: https://tinyurl.com/42fpndyw [accessed 2021-11-30]
31. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]




XSL•FO
RenderX

32. Nachega JB, Leisegang R, Kallay O, Mills EJ, Zumla A, Lester RT. Mobile health technology for enhancing the COVID-19 response in Africa: a potential game changer? Am J Trop Med Hyg 2020 Jul;103(1):3-5 [FREE Full text] [doi: 10.4269/ajtmh.20-0506] [Medline: 32476643]
33. Kanaabi M. 42% of Uganda's population now connected to internet. New Vision. 2020 May 30. URL: https://www.newvision.co.ug/news/1517233/-uganda-population-connected-internet [accessed 2021-11-30]

## Abbreviations

**EWAS:** Early Warning System for In-Hospital Transfer of COVID-19
**MakRIF:** Makerere University Research Innovation Fund
**mHealth:** mobile health
**PCR:** polymerase chain reaction
**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 27.01.21; peer-reviewed by D Pförringer, L McCann; comments to author 20.03.21; revised version received 15.04.21; accepted 31.05.21; published 17.12.21.*

*Please cite as:*
*Mbiine R, Nakanwagi C, Lekuya HM, Aine J, Hakim K, Nabunya L, Tomusange H*
*An Early Warning Mobile Health Screening and Risk Scoring App for Preventing In-Hospital Transmission of COVID-19 by Health Care Workers: Development and Feasibility Study*
*JMIR Form Res 2021;5(12):e27521*
*URL: https://formative.jmir.org/2021/12/e27521*
*doi:10.2196/27521*
*PMID:34793321*

©Ronald Mbiine, Cephas Nakanwagi, Herve Monka Lekuya, Joan Aine, Kawesi Hakim, Lilian Nabunya, Henry Tomusange. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Examining the Impact of Digital Components Across Different Phases of Treatment in a Blended Care Cognitive Behavioral Therapy Intervention for Depression and Anxiety: Pragmatic Retrospective Study

Monica S Wu[1*], PhD; Robert E Wickham[2,3*], PhD; Shih-Yin Chen[1], MS; Connie Chen[1], MD; Anita Lungu[1], PhD

[1]Lyra Health, Burlingame, CA, United States

[2]Department of Psychological Sciences, Northern Arizona University, Flagstaff, AZ, United States

[3]Department of Psychology, Palo Alto University, Palo Alto, CA, United States

[*]these authors contributed equally

**Corresponding Author:**
Monica S Wu, PhD
Lyra Health
287 Lorton Ave
Burlingame, CA, 94010
United States
Phone: 1 877 505 7147
Email: mwu@lyrahealth.com

## Abstract

**Background:** Depression and anxiety incur significant personal and societal costs. Effective psychotherapies exist, such as cognitive behavioral therapy (CBT); however, timely access to quality care is limited by myriad barriers. Blended care therapy models incorporate traditional face-to-face therapy with scalable, digital components of care, expanding the reach of evidence-based care.

**Objective:** The aim of this study is to determine the effectiveness of a blended care CBT program (BC-CBT) in real-world settings and examine the unique impacts of the (1) digital components of care (video lessons and digital exercises) and (2) phase of treatment (early versus late) in decreasing symptoms of anxiety and depression.

**Methods:** This retrospective cohort analysis included 3401 US-based individuals enrolled in a BC-CBT program, who presented with clinical levels of depression and/or anxiety. The treatment program consisted of regular therapy sessions augmented by clinician-assigned digital video lessons and exercises. A growth curve model incorporating time-varying covariates examined the relationship between engagement with BCT components (ie, therapy sessions, digital video lessons, and digital exercises) during the early (weeks 0-7) and late (weeks 8-15) phases of treatment, and weekly symptom reports on depression and anxiety measures.

**Results:** On average, a significant decline in depression and anxiety symptoms was observed during the initial weeks of treatment ($P<.001$), with a continued, though slower, decline over subsequent weeks ($P<.001$). Each session completed was associated with significant decreases in anxiety (b=–0.72) and depression (b=–0.83) in the early phase, as well as in the late phase (anxiety, b=–0.47; depression, b=–0.27). Significant decreases in anxiety (b=–0.15) and depression (b=–0.12) were observed for time spent on video lessons (measured in 10-minute intervals) in the early phase of treatment. Engaging with exercises was associated with statistically significant increases in anxiety symptoms (b=0.03) during the early phase of treatment. However, sensitivity analyses examining the effects of exercises in isolation revealed significant decreases in anxiety (b=–0.05) in the early phase, suggesting a potential suppression effect in the larger model.

**Conclusions:** Using a retrospective cohort design, therapy sessions and digital video lessons were uniquely predictive of improvements in depression and anxiety symptoms, and their effects were modulated based on the phase of treatment (early vs late). Future research should investigate whether other treatment variables, such as therapeutic alliance or familiarity with technology, are related to differential effects on various components of care.

***(JMIR Form Res 2021;5(12):e33452)*** doi:10.2196/33452



XSL•FO
RenderX

**KEYWORDS**

blended care; cognitive-behavioral therapy; depression; anxiety; digital; phase; mental health; digital health; digital therapy

## Introduction

Anxiety and depression are leading causes of disability worldwide, costing the global economy US $1 trillion each year due to lost productivity [1]. Given these significant societal and personal burdens [2,3], access to timely, quality care is imperative. Cognitive behavioral therapy (CBT) is considered the gold-standard psychotherapy for anxiety and depression and has robust empirical support [4]. However, obtaining access to CBT is hindered by various barriers, including the costs of therapy, long waiting lists, paucity of qualified mental health professionals, and lack of mental health resources (especially in rural areas) [5].

To overcome these barriers, teletherapy and internet-based CBT (iCBT) are promising ways to bridge the gap in accessing quality mental health care through technology. Video-based teletherapy allows treatment sessions to occur over a videoconferencing platform in real time, and it has shown similarly robust treatment outcomes when compared to in-person CBT for depression and anxiety [6]. Teletherapy facilitates access to therapy regardless of travel limitations or lack of resources in certain geographic locations, expanding the reach of evidence-based care [7].

Similarly, the use of iCBT permits remote access to evidence-based therapy materials, delivered through various digital media (text, audio, video, interactive, and gamified formats) [8]. Treatment materials are deliverable in a scalable and asynchronous manner, allowing for broader dissemination and convenient access to evidence-based therapy content. Targeting anxiety and depression through iCBT has shown significant decreases in symptom severity [9,10], although greater treatment effects are observed when iCBT is performed with therapist support versus without [11]. Indeed, difficulties with retention and smaller treatment effects arise without therapist guidance [12], especially when implemented in real word conditions [13], highlighting the potency of therapist involvement. Combining teletherapy and iCBT harnesses the advantages of both modalities of treatment, lending support for a blended care model of therapy.

Blended care therapy combines traditional face-to-face therapy with digital components of care that cover key therapeutic concepts and skills. Through therapist involvement in regular therapy sessions, clients are able to receive personalized care, discuss therapy content in depth, and be held accountable for homework completion. The digital materials help maintain fidelity to evidence-based practices, reinforce key concepts and skills outside of session, and facilitate rapid dissemination of therapeutic content in a scalable manner [14]. Blended care CBT (BC-CBT) integrates the benefits of both modalities and is a potentially cost-effective manner of delivering treatment [15]. Additionally, CBT delivered in a blended care format has preliminarily demonstrated significant decreases in depression and anxiety symptoms, highlighting its effectiveness in real-world settings [16].

Despite these benefits, there is limited research on the effectiveness of BC-CBT at scale, and whether the different digital components of care (eg, video lessons, digital exercises) have varied impacts on treatment outcome. Teasing apart the unique contributions of the digital components will help better inform which components are most potent in decreasing symptom severity. Additionally, the timing of when various care components are assigned could be an influential factor. Given that symptom changes in psychotherapy often do not follow a linear path [17], it is prudent to consider how different treatment components may contribute to symptom reduction in the early versus late phases of therapy. As various treatment components and process variables may be more or less potent at different times [18], examining these variables in a time-sensitive manner will allow for increased precision and personalization of care to optimize treatment outcomes. Early response and treatment gains at the beginning of therapy have received particular attention [19], as clients who experience early response to treatment tend to achieve better posttreatment outcomes for anxiety and depression [20]. The later phase of therapy is critically important as well, as it typically focuses on translating the therapeutic concepts and skills (learned in the earlier phase of therapy) into day-to-day practice, which is crucial for generalizing these skills [21]. Indeed, facing challenging situations and implementing regular out-of-session skills practices lead to long-term decreases in symptom severity and strengthen the durability of treatment gains [22,23]. Taken together, these findings highlight the importance of examining symptom changes in the early versus late phases of therapy, given the unique impact of each phase of therapy.

Ultimately, a gap remains in our understanding of the unique impact of the digital components of care in BC-CBT, as well as how the phase of treatment may impact treatment outcomes. Consequently, this study aims to evaluate the unique impact of treatment sessions, digital video lessons, and digital exercises in a blended care CBT intervention for depression and anxiety in real-world settings, examined across early versus late phases of treatment. These data are intended to better inform how to optimize care by customizing the different treatment components within a BC-CBT program and the timing of their assignment.

## Methods

### Study Design and Procedures

This study employed a retrospective cohort design, analyzing data collected as part of routine quality control for the BC-CBT program at Lyra Health. Lyra Health is partnered with Lyra Clinical Associates to provide therapeutic services to clients through the BC-CBT program. Clients are employees (or their dependents) of companies that offer mental health benefits through Lyra. Individuals were informed of their behavioral health benefits through their employers, and they could receive care with no cost to them using a set number of Employee Assistance Program sessions or, for certain employers, could receive care through their employer-sponsored health plan.



XSL•FO
RenderX

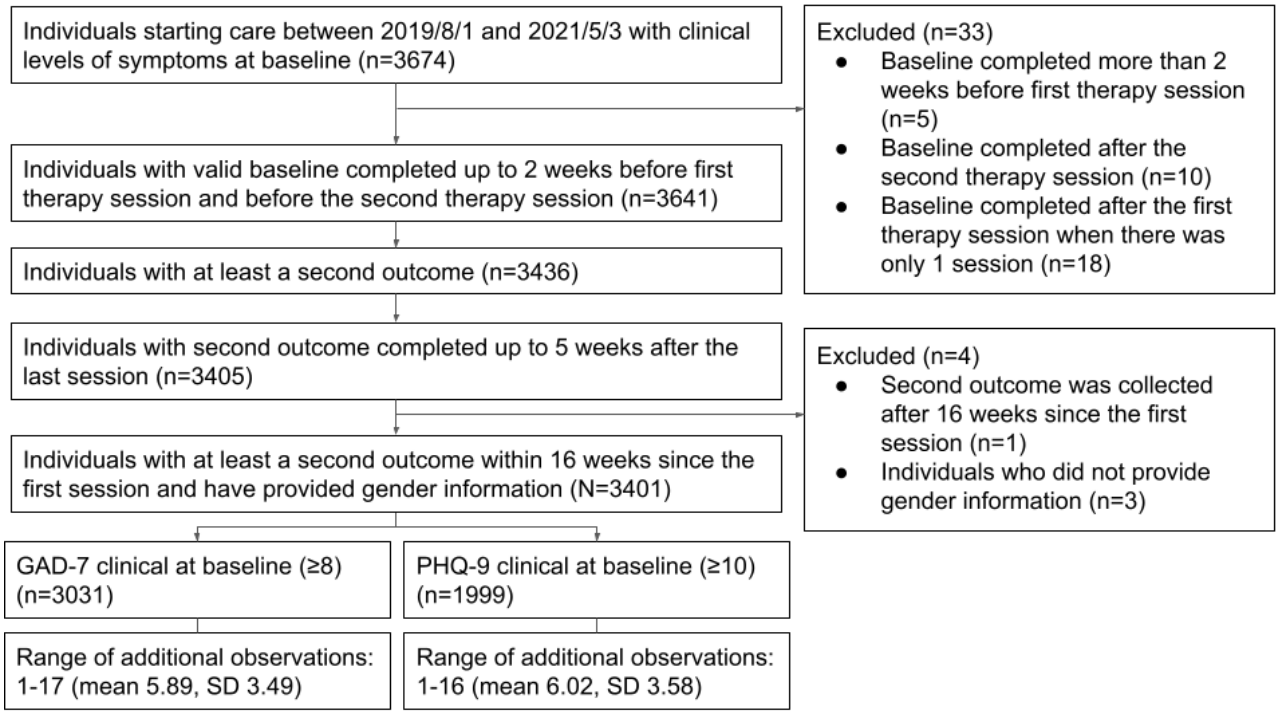
Individuals who accessed the services through their employer-sponsored health plan may have been required to pay a copay, coinsurance, or applicable deductible for sessions through the health plan. Interested clients registered on the web and completed a triage flow to indicate their presenting issues, preference for modality of care, and brief treatment history. If they were appropriate for the BC-CBT program, they were eligible to select the program on the web and book care with a provider. All clinical activities (therapy sessions, questionnaires, digital activities) were conducted over a secure, Health Insurance Portability and Accountability Act (HIPAA)–compliant online platform developed by Lyra Health. Clients were prompted to complete standardized measures of anxiety and depression weekly for the duration of care. Clients were assigned digital activities by their providers, and engagement with and completion of these activities were tracked through the online platform. This retrospective analysis of deidentified data collected during therapy was determined to not be human subjects research by the Palo Alto University institutional review board.

## Participants and Data Inclusion

Participants included in the study were individuals who started BC-CBT treatment between August 1, 2019, and May 3, 2021. Participants were required to have scored above the clinical cutoff for either the Generalized Anxiety Disorder-7 (GAD-7; score ≥8) or the Patient Health Questionnaire-9 (PHQ-9; score ≥10) on a valid baseline assessment (N=3674). To meet inclusion criteria for the BC-CBT program, clients were required to be 18 years of age or older and to be willing to see a provider via video. Clients were excluded from the BC-CBT program if they reported active suicidality, self-harm, or homicidality, or if they had a current diagnosis of severe alcohol/substance use disorder(s), psychiatric disorder with psychotic features that are not stabilized by medications, or unstable bipolar disorder.

Baseline assessments were considered invalid if they were collected more than 2 weeks prior to the first therapy session or after the second therapy session with the provider. Additionally, participants were considered to be missing a valid second assessment if no additional assessment beyond the baseline was completed within 5 weeks after the last therapy session. Assessments were also excluded if they were collected more than 16 weeks after the first therapy session, which represents the mean plus 1 standard deviation of the treatment duration for the sample. Please see Figure 1 for a comprehensive diagram of the participant flow.

**Figure 1.** Participant flow. GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9.




## Blended Care Therapy Program

The BC-CBT program included both regular face-to-face therapy sessions with the providers and digital components of care. Providers conducted therapy sessions via video-based teletherapy, generally starting with weekly sessions and gradually titrating to biweekly sessions. Providers assigned digital activities through the platform, and these activities were personalized based on the client's presenting issues. Completion of assigned digital activities could be monitored in real time via the platform by providers, who were encouraged to send secure messages to their client to remind and/or reinforce completion of assignments. More details regarding the BC-CBT online platform can be accessed via Lungu et al (2020) [16].

### *Therapy Sessions*

A short-term, goal-based model was employed, and providers used evidence-based care based on principles from CBT, dialectical behavior therapy (DBT), and acceptance and commitment therapy (ACT). Providers were extensively vetted



XSL•FO
RenderX

through a rigorous multistep interview process, and received intensive training in CBT and the proprietary online platform. Ongoing quality assurance was conducted via random session video reviews, regular consultation meetings, and continuing education presentations.

### *Digital Activities*

The digital components of the BCT program included *digital video lessons* and *digital exercises*. *Digital video lessons* employ a storytelling approach, which has been shown to be helpful for enhancing relatability and normalization of the presenting issues [24]. These videos follow a character going through therapy, presenting key CBT-based concepts and skills. Example topics include managing emotions, addressing thinking traps, challenging avoidance, and mindful awareness. The video lessons are ~8 minutes in length on average, and a brief quiz is administered at the end of each video lesson to check for understanding.

*Digital exercises* are analogous to paper logs or handouts that facilitate practice of therapy skills. They include awareness-building exercises (eg, thought record) or practice-oriented exercises (eg, exposure practice, behavioral activation, distress tolerance). Providers assign digital exercises to the clients based on their presenting needs, and clients are able to complete them asynchronously at their convenience. Exercise responses can be viewed in real time through the platform by both the provider and the client, and the provider is able to comment on the exercise to reinforce completion and troubleshoot any issues that arise.

## Measures

### *Demographics*

Information regarding client demographics (eg, the client's sex, ethnicity, and birthdate) was collected through the initial intake form completed on the online platform. Minority status was defined by the selection of a non-White group, including American Indian or Alaska Native, Asian or Pacific Islander, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, "multiple," or "other."

### *PHQ-9 Score*

The Patient Health Questionnaire 9 (PHQ-9) is a 9-item self-report questionnaire that assesses the presence and severity of depressive symptoms in the past week. A cutoff score of ≥10 on the PHQ-9 has been utilized to indicate a likely diagnosis of major depression [25].

### *GAD-7 Score*

The GAD-7 is a 7-item self-report questionnaire that assesses the presence and severity of anxiety symptoms in the past week [26]. A cutoff score on the GAD-7 of ≥8 has been used to indicate a likely diagnosis of generalized anxiety disorder [27].

Both symptom severity measures have undergone extensive validation in numerous clinical trials, demonstrating strong psychometric properties as supported by high validity, reliability, and treatment sensitivity [28].

### *Digital Activity Engagement*

The online platform records whether or not the client has engaged with the digital video lesson or the digital exercise. The time spent watching digital video lessons is recorded in the system and was coded in 10-minute intervals for the purposes of this study. The completion of digital exercises is logged in the system each time it is submitted by the client.

## Data Analyses

A mixed effects modeling approach to growth curve analysis was employed, which accommodates individually varying time intervals for outcome responses while accounting for missing data under the conditional missing at random (MAR) assumption. This approach also allows for the incorporation of time-invariant covariates (TICs) to model the impact of stable client-level attributes on individual trajectories, as well as time-varying covariates (TVCs) to model the predictive utility of attributes that vary across clients during the course of treatment. Analyses were conducted with SAS PROC MIXED, version 9.4, using restricted maximum likelihood estimation [29]. Because the focus of therapy is tailored and different outcomes may be observed based on the primary presenting issue [30], analyses were conducted separately for clients meeting the threshold for elevated anxiety (GAD-7 ≥8) or depression (PHQ-9 ≥10). If a client exceeded the thresholds for both anxiety and depression, they were included in both analyses.

Following cleaning and validation, a series of hierarchically structured individual growth curve models were examined. The initial analysis (Model 1) for each outcome (PHQ-9, GAD-7) featured a conditional model characterized by fixed effects corresponding to a second-order (quadratic) trajectory, along with client-level random effects for the intercept and linear trajectory components and a provider-level random effect for the intercept. The initial model also included the number of therapy sessions and digital exercises completed during the past 7 days, as well as the amount of time spent engaged with digital lessons (in 10-minute intervals) as TVCs.

Next, in Model 2, we incorporated a dummy-coded variable indicating the phase of treatment (late phase: Weeks 0-7=0; Weeks 8-15=1) along with the interaction terms between the phase indicator and the engagement TVCs. This allowed the magnitude of TVC effects to differ across the early and late phases of treatment; i.e., the inclusion of the late phase*TVC interaction terms means that the first-order coefficients for therapy sessions, digital lessons, and exercises represent the simple effects of these predictors during the early phase of the study period [31]. In addition, the interaction terms describe the change in the magnitude of each TVC when considered during the late phase relative to the early phase.

In Model 3, several demographic variables were added as TICs to examine the impact of age, gender, and race on individual trajectories. Finally, several sensitivity analyses were conducted to evaluate the impact of receiving a therapy session and engaging with digital lessons during the same week, along with the isolated association between digital exercise engagement and weekly symptoms.



XSL•FO
RenderX

## Results

### Participant Characteristics

Table 1 includes the participants' demographic and baseline characteristics for the entire sample and each of the analysis samples (GAD-7 and PHQ-9), as well as the engagement in therapy sessions, digital lessons, and exercises during the first 16 weeks.

Figures 2 and 3 are bar charts that respectively illustrate the GAD-7 and PHQ-9 mean scores over the study period. The length of the error bars represents the standard error of the mean. To interpret the "Week" values, Week=–1 is the week before the first session, and Week=1 is the first week after the first session.

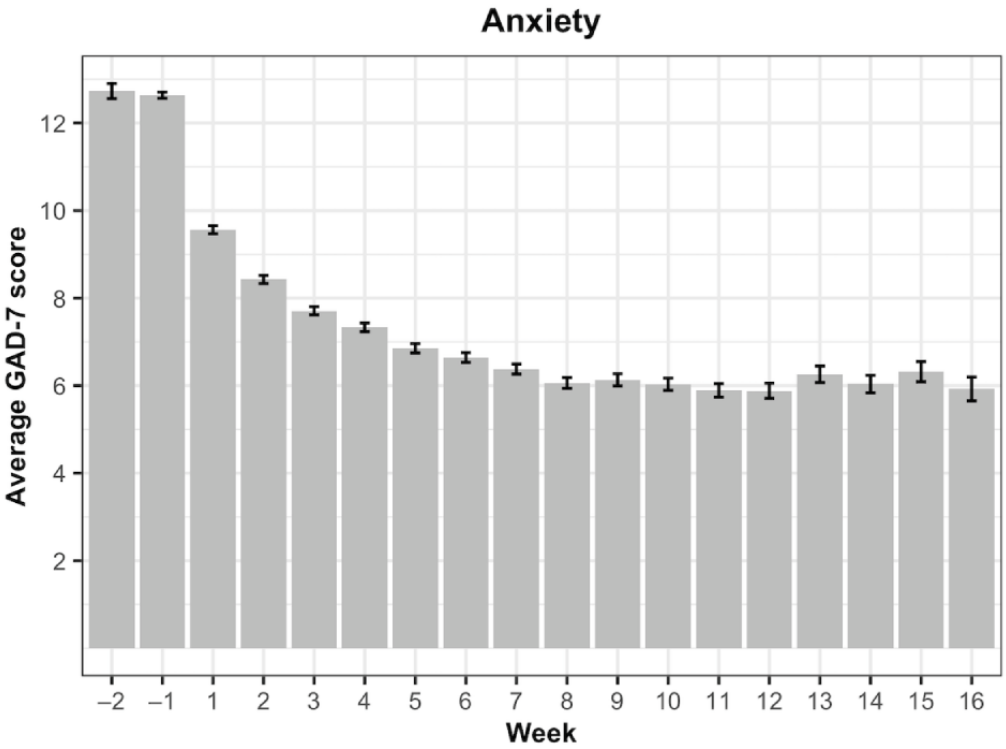
**Table 1.** Demographic information and engagement in treatment components.

| | Entire sample (N=3401) | GAD-7[a] sample (n=3031) | PHQ-9[b] sample (n=1999) |
|---|---|---|---|
| Age, mean (SD) | 33.15 (8.68) | 33.22 (8.66) | 33.16 (8.82) |
| Female sex, n (%) | 2218 (65.22) | 1987 (65.56) | 1304 (65.23) |
| **Race/ethnicity, n (%)** | | | |
| Minority | 1731 (50.90) | 1547 (51.04) | 1043 (52.18) |
| Unknown | 136 (4.00) | 121 (3.99) | 95 (4.75) |
| Baseline GAD-7 score, mean (SD) | 11.83 (4.08) | 12.64 (3.52) | 12.14 (4.68) |
| Baseline PHQ-9 score, mean (SD) | 10.72 (4.92) | 10.49 (5.08) | 13.99 (3.43) |
| Number of sessions completed, mean (SD) | 5.82 (3.05) | 5.84 (3.09) | 5.97 (3.14) |
| **Total number of out-of-session engagements, mean (SD)** | 43.36 (28.34) | 43.39 (28.45) | 44.80 (29.49) |
| Number of lessons completed, mean (SD) | 5.56 (3.54) | 5.53 (3.52) | 5.67 (3.65) |
| Number of exercises completed, mean (SD) | 10.56 (12.16) | 10.49 (12.10) | 10.96 (12.57) |
| Duration of care (weeks), mean (SD) | 6.97 (4.71) | 6.98 (4.72) | 7.13 (4.79) |

[a]GAD-7: Generalized Anxiety Disorder-7.

[b]PHQ-9: Patient Health Questionnaire-9.

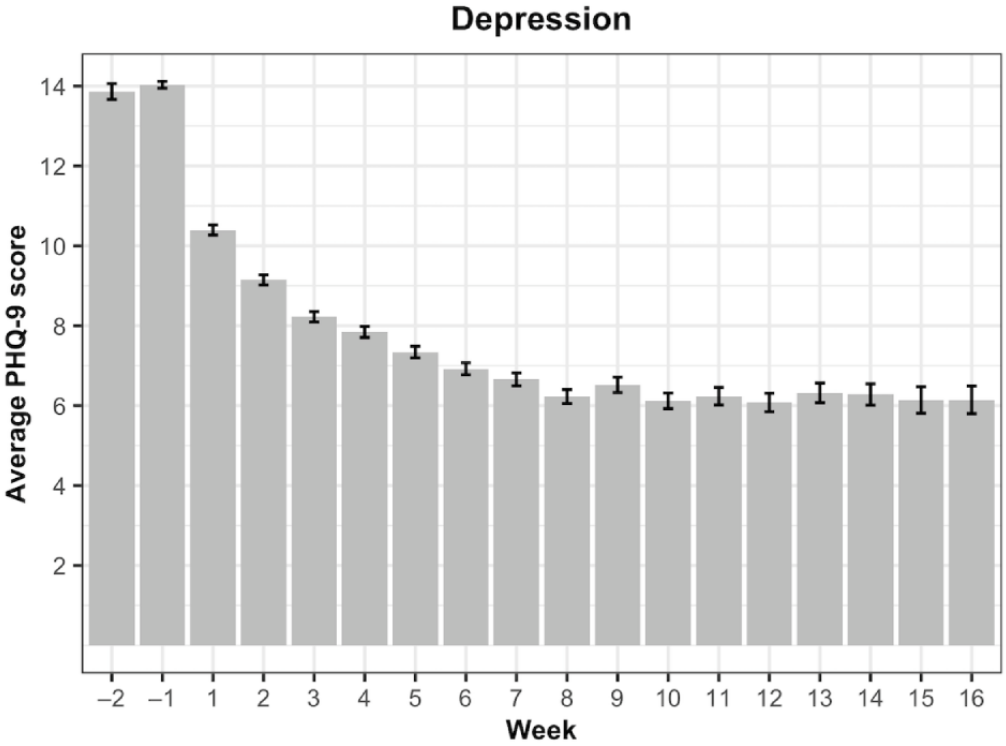
**Figure 2.** Mean Generalized Anxiety Disorder-7 (GAD-7) scores by week.






XSL•FO
RenderX

**Figure 3.** Mean Patient Health Questionnaire-9 (PHQ-9) scores by week.




## GAD-7 Score

Model 1 featured a linear and a quadratic trajectory and time-varying indicators of engagement (therapy sessions, digital lessons, and digital exercises). The fixed coefficient for the first-order effect of the treatment week revealed that on average, GAD-7 symptom scores decreased considerably during the first week of treatment (b=–1.09, 95% CI –1.12 to –1.06; *P*<.001). However, a significant quadratic coefficient (b=0.05, 95% CI 0.05-0.05; *P*<.001) suggests that the rate of decline diminishes over time. Fixed effect coefficients for the TVCs revealed that clients who attended a therapy session during the last 7 days reported significantly lower GAD-7 scores (b=–0.68, 95% CI –0.77 to –0.59; *P*<.001), and that every 10 minutes of digital lessons completed was associated with significantly lower GAD-7 scores (b=–0.14, 95% CI –0.19 to –0.08; *P*<.001). Contrary to our expectations, engagement with digital exercises was associated with a slight increase in GAD-7 scores (b=0.03, 95% CI 0.00-0.06; *P*=.045).

In Model 2, the TVC coefficients were allowed to differ across the phases of treatment. The first-order effect of the phase indicator variable (late phase) suggests that the average anxiety trajectory shifts upwards slightly during the later phase of treatment (b=0.44, 95% CI 0.20-0.68; *P*<.001), relative to the early phase. A significant late phase*therapy sessions interaction emerged (b=0.25, 95% CI 0.03-0.47; *P*=.03), suggesting that the beneficial effect of a session became weaker during the late phase of treatment. A similar pattern emerged for digital lessons, such that the beneficial effect of completing 10 minutes of lessons was significantly weaker (b=0.22, 95% CI 0.07-0.36; *P*=.004) during the late phase of treatment. In contrast, the phase interaction involving digital exercises failed to reach significance.

Model 3 contained the same configuration of TVCs and phase interactions as Model 2, but incorporated several client-level TICs examining the impact of demographic variables. Although no differences in initial GAD-7 scores emerged as a function of age or race, females reported slightly higher baseline anxiety scores (b=0.37, 95% CI 0.15-0.60; *P*=.001) relative to male clients. The coefficients describing TVC effects across early and late phases were identical to Model 2. In sum, anxiety symptoms declined significantly over the course of treatment overall. Additionally, engagement with weekly therapy sessions and digital lessons was associated with lower anxiety symptoms, particularly during the early phase of treatment. Table 2 includes all point estimates and 95% confidence intervals for fixed effects for the GAD-7 analyses, and Table 3 shows the deviance and selection criteria.



XSL•FO
RenderX

**Table 2.** Generalized Anxiety Disorder-7 analysis results (n=3031).

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate (95% CI) | $t_{obs}$[a] (*df*) | *P* value | Estimate (95% CI) | $t_{obs}$ (*df*) | *P* value | Estimate (95% CI) | $t_{obs}$ (*df*) | *P* value |
| Intercept | 11.31 (11.17 to 11.46) | 153.80 (297) | <.001 | 11.36 (11.22 to 11.51) | 153.62 (307) | <.001 | 11.16 (10.93 to 11.40) | 91.87 (1485) | <.001 |
| Week | –1.09 (–1.12 to –1.06) | 65.75 (15,000) | <.001 | –1.09 (–1.13 to –1.06) | 65.89 (15,000) | <.001 | –1.09 (–1.13 to –1.06) | 65.91 (15,000) | <.001 |
| $Week^2$ | 0.05 (0.05 to 0.05) | 36.12 (16,000) | <.001 | 0.04 (0.04 to 0.05) | 29.81 (17,000) | <.001 | 0.04 (0.04 to 0.05) | 29.82 (17,000) | <.001 |
| Therapy sessions in the last 7 days | –0.68 (–0.77 to –0.59) | 14.36 (18,000) | <.001 | –0.72 (–0.82 to –0.61) | 13.58 (18,000) | <.001 | –0.72 (–0.82 to –0.61) | 13.59 (18,000) | <.001 |
| Digital exercises in the last 7 days | 0.03 (0.00 to 0.06) | 2.01 (19,000) | .04 | 0.03 (0.00 to 0.06) | 2.12 (19,000) | .03 | 0.03 (0.00 to 0.06) | 2.10 (19,000) | .04 |
| Digital lessons in the last 7 days | –0.14 (–0.19 to –0.08) | 4.74 (18,000) | <.001 | –0.15 (–0.21 to –0.09) | 4.91 (18,000) | <.001 | –0.15 (–0.21 to –0.09) | 4.86 (18,000) | <.001 |
| Late phase | __[b] | — | — | 0.44 (0.20 to 0.68) | 3.60 (18,000) | <.001 | 0.44 (0.20 to 0.68) | 3.61 (18,000) | <.001 |
| Late phase*therapy sessions in the last 7 days | — | — | — | 0.25 (0.03 to 0.47) | 2.19 (18,000) | .03 | 0.25 (0.03 to 0.47) | 2.20 (18,000) | .03 |
| Late phase*digital exercises in the last 7 days | — | — | — | 0.04 (–0.03, to 0.11) | 1.06 (18,000) | .29 | 0.04 (–0.03 to 0.11) | 1.07 (18,000) | .29 |
| Late phase*digital lessons in the last 7 days | — | — | — | 0.22 (0.07 to 0.36) | 2.88 (18,000) | .004 | 0.21 (0.07 to 0.36) | 2.85 (18,000) | .004 |
| Age | — | — | — | — | — | — | 0.00 (–0.01 to 0.01) | 0.27 (2941) | .79 |
| Gender | — | — | — | — | — | — | 0.37 (0.15 to 0.60) | 3.24 (2986) | .001 |
| Minority ethnicity | — | — | — | — | — | — | –0.13 (–0.35 to 0.09) | 1.15 (2984) | .25 |
| No ethnicity reported | — | — | — | — | — | — | 0.47 (–0.09 to 1.02) | 1.64 (3016) | .10 |

[a]$t_{obs}$: observed *t* test result.

[b]—: not applicable (variables not entered into model).

**Table 3.** Deviance and selection criteria for the 3 models in the Generalized Anxiety Disorder-7 analysis.

| | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Deviance (–2LL[a]) | 111676.4 | 111622.0 | 111620.8 |
| Akaike information criterion | 111686.4 | 111632.0 | 111630.8 |
| Bayesian information criterion | 111704.2 | 111649.9 | 111648.7 |

[a]—2LL: log-likelihood ratio.

## PHQ-9 Score

As before, the fixed coefficient for the first-order effect of treatment week for Model 1 revealed that on average, PHQ-9 scores decreased significantly during the first week of treatment (b=–1.24, 95% CI –1.28 to –1.20; *P*<.001). A significant quadratic coefficient (b=0.05, 95% CI 0.05-0.06, *P*<.001) indicates that the rate of decline diminishes over the course of treatment. Fixed effect coefficients for the TVCs revealed that clients who attended a therapy session during the last 7 days



XSL•FO
RenderX

reported significantly lower PHQ-9 scores (b=–0.73, 95% CI –0.85 to –0.61; *P*<.001). In addition, completion of digital exercises (b=–0.04, 95% CI –0.08 to 0.00; *P*=.03) and digital lessons (b=–0.01, 95% CI –0.17 to –0.03; *P*=.008) over the prior week was associated with significantly lower PHQ-9 scores.

In Model 2, the first-order effect of late phase suggests that the average depression trajectory shifts upwards slightly during Weeks 8 to 15 of treatment (b=0.38, 95% CI 0.08-0.69; *P*=.01). A significant late phase*therapy sessions interaction emerged (b=0.56, 95% CI 0.27-0.84; *P*<.001), suggesting that the beneficial effect of a session became weaker during the late phase of treatment. A similar pattern emerged for digital lessons, such that the beneficial effect of completing 10 minutes of lessons was significantly weaker (b=0.28, 95% CI 0.09-0.46; *P*=.003) during the late phase of treatment. In contrast, the phase interaction involving digital exercises failed to reach significance.

For Model 3, no differences in initial PHQ-9 scores emerged as a function of age, race, or gender. Overall, depression symptoms declined significantly over the course of treatment. Additionally, engagement with all program elements was associated with lower depression symptoms, with therapy sessions and digital lessons exhibiting a stronger effect during the early phase of treatment. Table 4 includes all point estimates and 95% confidence intervals for fixed effects for the PHQ-9 analyses, and Table 5 shows the deviance and selection criteria.

**Table 4.** Patient Health Questionnaire-9 analysis results (n=1999).

| | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate (95% CI) | $t_{obs}$[a] (*df*) | *P* value | Estimate (95% CI) | $t_{obs}$ (*df*) | *P* value | Estimate (95% CI) | $t_{obs}$ (*df*) | *P* value |
| Intercept | 12.43 (12.24 to 12.62) | 130.55 (276) | <.001 | 12.51 (12.32 to 12.70) | 130.54 (284) | <.001 | 12.38 (12.06 to 12.71) | 74.43 (1247) | <.001 |
| Week | –1.24 (–1.28 to –1.20) | 56.79 (9682) | <.001 | –1.24 (–1.28 to –1.20) | 56.89 (9714) | <.001 | –1.24 (–1.28 to –1.20) | 56.88 (9717) | <.001 |
| Week$^2$ | 0.05 (0.05 to 0.06) | 30.74 (11,000) | <.001 | 0.05 (0.04 to 0.05) | 25.51 (12,000) | <.001 | 0.05 (0.04 to 0.05) | 25.51 (12,000) | <.001 |
| Therapy sessions in the last 7 days | –0.73 (–0.85 to –0.61) | 11.87 (12,000) | <.001 | –0.82 (–0.96 to –0.69) | 12.07 (12,000) | <.001 | –0.83 (–0.96 to –0.69) | 12.08 (12,000) | <.001 |
| Digital exercises in the last 7 days | –0.04 (–0.08 to 0.00) | 2.18 (13,000) | .03 | –0.03 (–0.07 to 0.01) | 1.60 (13,000) | .11 | –0.03 (–0.07 to 0.01) | 1.59 (13,000) | .11 |
| Digital lessons in the last 7 days | –0.10 (–0.17 to –0.03) | 2.67 (12,000) | .01 | –0.12 (–0.19 to –0.04) | 2.95 (12,000) | .003 | –0.12 (–0.19 to –0.04) | 2.94 (12,000) | .003 |
| Late phase | —[b] | — | — | 0.38 (0.08 to 0.69) | 2.45 (12,000) | .01 | 0.38 (0.08 to 0.69) | 2.45 (12,000) | .01 |
| Late phase*therapy sessions in the last 7 days | — | — | — | 0.56 (0.27 to 0.84) | 3.82 (12,000) | <.001 | 0.56 (0.27 to 0.84) | 3.83 (12,000) | <.001 |
| Late phase*digital exercises in the last 7 days | — | — | — | 0.02 (–0.07 to 0.11) | 0.43 (12,000) | .67 | 0.02 (–0.07 to 0.11) | 0.43 (12,000) | .67 |
| Late phase*digital lessons in the last 7 days | — | — | — | 0.28 (0.09 to 0.46) | 2.96 (12,000) | .003 | 0.28 (0.09 to 0.46) | 2.96 (12,000) | .003 |
| Age | — | — | — | — | — | — | 0.01 (–0.01 to 0.02) | 0.61 (1986) | .54 |
| Gender | — | — | — | — | — | — | 0.23 (–0.09 to 0.54) | 1.41 (1991) | .16 |
| Minority ethnicity | — | — | — | — | — | — | –0.07 (–0.38 to 0.24) | 0.43 (1998) | .67 |
| No ethnicity reported | — | — | — | — | — | — | 0.28 (–0.45 to 1.00) | 0.75 (2058) | .45 |

[a]$t_{obs}$: observed *t* test result.

[b]—: not applicable (variables not entered into model).



XSL•FO
RenderX

**Table 5.** Deviance and selection criteria for the 3 models in the Patient Health Questionnaire-9 analysis.

| | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| Deviance (–2LL[a]) | 76960.5 | 76905.2 | 76913.4 |
| Akaike information criterion | 76970.5 | 76915.2 | 76923.4 |
| Bayesian information criterion | 76988.2 | 76932.8 | 76941.1 |

[a]—2LL: log-likelihood ratio.

### Sensitivity Analysis

To investigate the extent to which the unexpected findings for the digital exercises in Model 1 of the GAD-7 analysis can be explained by multicollinearity or suppression effects, we specified a series of follow-up models examining digital exercises in isolation. A significant effect emerged, such that completing a digital exercise was associated with an expected b=–0.05 unit decrease in GAD-7 scores (95% CI –0.08 to –0.02; $P<.001$).

## Discussion

Findings from this study revealed significant declines in anxiety and depression symptoms, with a steep initial decline that became less pronounced over time. The initial model showed that engagement with therapy sessions and digital lessons during the prior 7 days was uniquely associated with lower anxiety and depression symptoms. Although engagement with digital exercises over the past week was associated with lower depression symptoms, a positive coefficient emerged in the analysis on anxiety symptoms. Sensitivity analyses revealed that the unexpected positive relationship between use of digital exercises and anxiety symptoms may be attributable to collinearity with therapy sessions and digital lessons. Analyses examining differences across the early (Weeks 0-7) and late (Weeks 8-15) phases of therapy found that for both anxiety and depression, therapy sessions and digital lessons (but not digital exercises) were significantly stronger (more negative) during the early phase of treatment. These effects remained after incorporating age, gender, and ethnicity.

Overall, significant declines in both anxiety and depression were observed throughout treatment, demonstrating the efficacy of BC-CBT in decreasing these symptomatologies. These robust effects lend support to a blended care model in addressing anxiety and depressive symptoms [16], and such a model can be a potentially more scalable and cost-effective method of providing evidence-based care compared to traditional methods of delivering psychotherapy [15]. Additionally, steeper declines in symptom severity were observed during the early stages of therapy, and therapy sessions and video lessons were significantly more potent in decreasing symptomatology during the early phase of treatment. Given that the initial stages of therapy are largely dedicated to providing pertinent psychoeducation and establishing key therapeutic concepts and skills, therapy sessions and digital video lessons appear to be effective ways to reinforce the uptake of this crucial information. Specifically, a key component of early psychoeducation is to establish the rationale for treatment and clarify how current cognitive and behavioral patterns contribute to the maintenance of symptomology. This crucial information has been shown to facilitate client reappraisals and contributes to lower avoidance of challenging situations early on in treatment [32], which is key for decreasing symptom severity and functional impairment. Consequently, providers should capitalize on the robust effects of these treatment components earlier on in therapy, ensuring that sessions are occurring consistently and that video lessons are being routinely assigned to clients in order to reinforce key therapeutic concepts. These findings also highlight the importance of setting positive expectations and obtaining client buy-in early on in therapy [33], given that the early phase of treatment often incites hope and a sense of agency [34]. This will optimally position clients to be their own enactors of change, motivating them to learn and practice coping skills later on in therapy.

Indeed, the later phase of therapy is primarily dedicated to the actual practice of skills in day-to-day life, which is challenging for clients as they start to practice these coping skills in more difficult, real-world situations. As such, it is not surprising to see a slower decline in symptoms in the latter stage of therapy, which is expected due to the increased difficulty of therapeutic tasks [17,35]. The later stages of therapy are critical to achieving reliable clinical improvement and durability of treatment gains, given the opportunities for real-world generalization of therapy skills. To facilitate these crucial steps, future studies would benefit from investigating how to increase the potency of digital tools and engagement in the later phase and further develop digital content that is better suited for later stages in treatment (eg, more in-depth skills review and practices). Ultimately, although the early and late phases of therapy have unique focuses, they both play crucial roles in achieving positive treatment outcomes.

Having a therapy session and watching digital video lessons were uniquely associated with decreases in anxiety and depression, highlighting the potency of both traditional face-to-face therapy and the digital components within a blended care model. Meeting with a therapist for regular sessions allows the client to receive individualized care, in-depth exploration of key concepts and skills, and increased accountability. Indeed, therapy sessions were relatively more impactful in decreasing symptomology, underscoring the importance of therapist-led treatment [36,37]. It is notable that completing digital video lessons also uniquely contributed to decreases in anxiety and depression, supporting its potential standalone utility. The digital video lessons help teach and reinforce key therapy concepts in an engaging manner, allow for asynchronous review of the materials at the client's convenience, and facilitate dissemination of therapeutic content at scale. These benefits are particularly important to consider when resources are limited and therapy



XSL•FO
RenderX

sessions cannot be conducted with regularity, which can be affected by financial barriers, difficulties obtaining reliable transportation to/from sessions, or lack of access to trained professionals [38,39]. In these cases, digital video lessons can be pivotal in expanding access to evidence-based treatment, given the support that digital interventions have received for effectively reducing access barriers [40].

There was an unexpected finding of digital exercises being associated with increased anxiety symptoms, but the effect size was relatively weak. Based on the sensitivity analyses, the follow-up models indicated a potential suppression effect for digital exercises; digital exercises in isolation contributed to decreases in anxiety and depressive symptoms. These findings are more consistent with extant literature, which reports a strong association between greater homework compliance and better treatment outcomes [41,42]. The relatively small effect sizes observed can be attributed to the measurement method of exercise completion. In fact, previous research has reported smaller effect sizes when homework completion is measured objectively and contemporaneously [41], both of which are typical with digital platforms. Additionally, clients may have generalized the skills to day-to-day life as therapy progressed and gradually decreased reliance on the digital exercise [21], causing them to practice the skills taught in the exercises but not actually log them in the platform, thereby diluting the strength of the association. Future studies should seek to further disentangle the overlap between various digital components of care and better establish the unique contributions through randomly assigned conditions and/or sequencing studies.

These findings should be considered within the context of several limitations and suggested future directions for research. First, this study employs a pragmatic retrospective design, limiting the ability to rule out regression to the mean effects. Future studies should seek to examine the efficacy of BC-CBT and its care components in a randomized controlled trial to instill further confidence in this model beyond a control condition. Within these limitations, it is important to note that the naturalistic design in real-world settings strengthens the ecological validity, and the growth curve model allows us to glean nuanced information about the temporality of changes and account for stable and varying client-level differences. Second, self-report measures were used to determine clinical outcomes. Although the measures used are extensively validated and possess strong psychometric properties, future studies should use a multimethod, multi-informant approach to measuring symptomatology to establish a more comprehensive evaluation [43]. Third, more general therapy process variables were not incorporated into the analyses. As such, future studies should seek to better understand the impact of these factors (eg, therapeutic alliance, treatment expectations, familiarity with technology, therapist reinforcement of skills practice) on engagement with digital activities and treatment outcome [44]. Fourth, the measurement of digital exercise completion could be improved in future studies. Although real-time data were gathered whenever clients completed exercises in the platform, data collection is incumbent upon the client electronically logging it, so it is possible that clients were practicing the coping skills noted in the digital exercise but not actually logging it in the system. Additionally, more nuanced data regarding the actual time clients spent on digital exercises may be helpful to examine in relation to outcomes. Fifth, given that the number of exercises assigned decreases significantly in the later phase of therapy, it is possible that there was less power to detect the impact of exercises on decreasing symptomology. Sixth, given that other types of digital engagement occurred in between sessions (eg, completion of assessments, messages between provider and client), it would also be helpful to examine their impact on treatment outcomes. Although they are inherently different from digital therapy content (ie, digital content directly teaching therapeutic concepts and skills), these other types of engagement may uniquely impact engagement in therapy overall. Seventh, this study focused on treatment outcomes during the acute phase of therapy. Future studies should seek to conduct long-term follow-ups to determine how robust BC-CBT is in maintaining treatment gains for anxiety and depression. Eighth, although anxiety and depression were the primary focuses of the study, clients may have had other presenting problems as well. Future studies should seek to examine the impact of comorbidities on treatment outcome. Ninth, this study focused on distinct outcomes (ie, anxiety and depression symptoms), and many clients met inclusion criteria for both conditions, resulting in their inclusion in both sets of analyses. Although this does not pose a meaningful threat to the validity of the analyses presented here, future studies drawing on larger samples should explore whether clients with comorbid presentations respond differently to the BC-CBT intervention than clients who only meet criteria for one condition. Tenth, providers anchored their interventions in CBT-based approaches and were able to implement a variety of interventions (eg, DBT, ACT). It would be of interest in future studies to investigate the potential impact of different approaches on outcomes.

Collectively, these findings highlight the robust effectiveness of a BC-CBT program in decreasing anxiety and depression in real-world settings. Given the myriad barriers to accessing evidence-based care, blended care programs combine the advantages of technology with traditional face-to-face therapy to facilitate greater dissemination of CBT. Timing of treatment components and the potency of certain digital activities (ie, video lessons) should be considered when implementing a blended care model. Ultimately, the unique contributions of various treatment components underscore the utility of a blended care model in effectively treating depression and anxiety.

## Acknowledgments

The authors would like to extend their gratitude for the Lyra Health blended care therapists, who are dedicated to helping their clients live more productive and fulfilling lives.



XSL•FO
RenderX

## Authors' Contributions

All authors were responsible for study conceptualization and for writing components of the original manuscript draft, as well as for reviewing and editing. SC was responsible for data curation and formal analysis. SC and REW were responsible for the formulation of the data analytic plan and data validation.

## Conflicts of Interest

MSW, SC, and AL, are employed by Lyra Health, receive income from Lyra Health, and have been granted equity in Lyra Health. CC is employed by Lyra Health and Lyra Clinical Associates, receives income from Lyra Health and Lyra Clinical Associates, and has been granted equity in Lyra Health. REW is a paid consultant for Lyra Health.

## References


1. The Lancet Global Health. Mental health matters. Lancet Glob Health 2020 Nov;8(11):e1352 [FREE Full text] [doi: 10.1016/S2214-109X(20)30432-0] [Medline: 33069297]
2. Wittchen HU. Generalized anxiety disorder: prevalence, burden, and cost to society. Depress Anxiety 2002;16(4):162-171 [FREE Full text] [doi: 10.1002/da.10065] [Medline: 12497648]
3. Kessler RC. The costs of depression. Psychiatr Clin North Am 2012 Mar;35(1):1-14 [FREE Full text] [doi: 10.1016/j.psc.2011.11.005] [Medline: 22370487]
4. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
5. Payne KA, Myhr G. Increasing access to cognitive-behavioural therapy (CBT) for the treatment of mental illness in Canada: a research framework and call for action. Healthc Policy 2010 Feb;5(3):e173-e185 [FREE Full text] [Medline: 21286263]
6. Fernandez E, Woldgabreal Y, Day A, Pham T, Gleich B, Aboujaoude E. Live psychotherapy by video versus in-person: a meta-analysis of efficacy and its relationship to types and targets of treatment. Clin Psychol Psychother 2021 Apr 07. [doi: 10.1002/cpp.2594] [Medline: 33826190]
7. Weightman M. Digital psychotherapy as an effective and timely treatment option for depression and anxiety disorders: implications for rural and remote practice. J Int Med Res 2020 Jun 12;48(6) [FREE Full text] [doi: 10.1177/0300060520928686] [Medline: 32527170]
8. Andersson G, Carlbring P, Ljótsson B, Hedman E. Guided internet-based CBT for common mental disorders. J Contemp Psychother 2013 May 21;43(4):223-233. [doi: 10.1007/s10879-013-9237-9]
9. Păsărelu CR, Andersson G, Nordgren LB, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
10. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
11. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
12. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
13. Baumel A, Edan S, Kane JM. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
14. Valerio Netto A. Application of blended care as a mechanism of action in the construction of digital therapeutics. Einstein (Sao Paulo) 2020;18:eMD5640 [FREE Full text] [doi: 10.31744/einstein_journal/2020MD5640] [Medline: 33237249]
15. Kooistra LC, Wiersma JE, Ruwaard J, Neijenhuijs K, Lokkerbol J, van Oppen P, et al. Cost and effectiveness of blended versus standard cognitive behavioral therapy for outpatients with depression in routine specialized mental health care: pilot randomized controlled trial. J Med Internet Res 2019 Oct 29;21(10):e14261 [FREE Full text] [doi: 10.2196/14261] [Medline: 31663855]
16. Lungu A, Jun JJ, Azarmanesh O, Leykin Y, Chen CE. Blended care-cognitive behavioral therapy for depression and anxiety in real-world settings: pragmatic retrospective study. J Med Internet Res 2020 Jul 06;22(7):e18723 [FREE Full text] [doi: 10.2196/18723] [Medline: 32628120]
17. Hayes AM, Laurenceau JP, Feldman G, Strauss JL, Cardaciotto L. Change is not always linear: the study of nonlinear and discontinuous patterns of change in psychotherapy. Clin Psychol Rev 2007;27(6):715-723 [FREE Full text] [doi: 10.1016/j.cpr.2007.01.008] [Medline: 17316941]
18. Gloster AT, Klotsche J, Gerlach AL, Hamm A, Ströhle A, Gauggel S, et al. Timing matters: change depends on the stage of treatment in cognitive behavioral therapy for panic disorder with agoraphobia. J Consult Clin Psychol 2014;82(1):141-153. [doi: 10.1037/a0034555] [Medline: 24188509]




XSL•FO
RenderX

19. Shalom JG, Aderka IM. A meta-analysis of sudden gains in psychotherapy: outcome and moderators. Clin Psychol Rev 2020;76:101827. [doi: 10.1016/j.cpr.2020.101827] [Medline: 32036232]
20. Beard JIL, Delgadillo J. Early response to psychological therapy as a predictor of depression and anxiety treatment outcomes: a systematic review and meta-analysis. Depress Anxiety 2019 Sep 24;36(9):866-878. [doi: 10.1002/da.22931] [Medline: 31233263]
21. Cummings C, Kazantzis N, Kendall PC. Facilitating homework and generalization of skills to the real world. In: Evidence-Based CBT for Anxiety and Depression in Children and Adolescents: A Competencies Based Approach. West Sussex, UK: John Wiley & Sons, Ltd; 2014:141-155.
22. Huppert JD, Roth Ledley D, Foa EB. The use of homework in behavior therapy for anxiety disorders. J Psychother Integr 2006;16(2):128-139. [doi: 10.1037/1053-0479.16.2.128]
23. Glenn D, Golinelli D, Rose RD, Roy-Byrne P, Stein MB, Sullivan G, et al. Who gets the most out of cognitive behavioral therapy for anxiety disorders? The role of treatment dose and patient engagement. J Consult Clin Psychol 2013 Aug;81(4):639-649 [FREE Full text] [doi: 10.1037/a0033403] [Medline: 23750465]
24. Williams AD, O'Moore K, Mason E, Andrews G. The effectiveness of internet cognitive behaviour therapy (iCBT) for social anxiety disorder across two routine practice pathways. Internet Interv 2014 Oct;1(4):225-229. [doi: 10.1016/j.invent.2014.11.001]
25. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
26. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
27. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
28. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
29. SAS Institute. SAS/STAT Documentation (Version 9.4). Cary, NC: SAS; 2014.
30. Flynn HA, Warren R. Using CBT effectively for treating depression and anxiety. Curr Psychiatry 2014;13(6):45-53 [FREE Full text]
31. Aiken LS, West SG. Multiple Regression: Testing and Interpreting Interactions. Newbury Park, CA: Sage; 1991.
32. Holtforth M, Grawe K, Castonguay LG. Predicting a reduction of avoidance motivation in psychotherapy: Toward the delineation of differential processes of change operating at different phases of treatment. Psychother Res 2006 Oct;16(5):639-644. [doi: 10.1080/10503300600608215]
33. DeFife JA, Hilsenroth MJ. Starting off on the right foot: common factor elements in early psychotherapy process. J Psychother Integr 2011;21(2):172-191 [FREE Full text] [doi: 10.1037/a0023889]
34. Howard KI, Lueger RJ, Maling MS, Martinovich Z. A phase model of psychotherapy outcome: causal mediation of change. J Consult Clin Psychol 1993 Aug;61(4):678-685. [doi: 10.1037//0022-006x.61.4.678] [Medline: 8370864]
35. Chu BC, Skriner LC, Zandberg LJ. Shape of change in cognitive behavioral therapy for youth anxiety: symptom trajectory and predictors of change. J Consult Clin Psychol 2013;81(4):573-587. [doi: 10.1037/a0033390] [Medline: 23796317]
36. Reardon ML, Cukrowicz KC, Reeves MD, Joiner TE. Duration and Regularity of Therapy Attendance as Predictors of Treatment Outcome in an Adult Outpatient Population. Psychother Res 2010;12(3):273-285. [doi: 10.1080/713664390]
37. Xiao H, Hayes JA, Castonguay LG, McAleavey AA, Locke BD. Therapist effects and the impacts of therapy nonattendance. Psychotherapy (Chic) 2017;54(1):58-65. [doi: 10.1037/pst0000103] [Medline: 28263652]
38. Patel V, Chowdhary N, Rahman A, Verdeli H. Improving access to psychological treatments: lessons from developing countries. Behav Res Ther 2011 Sep;49(9):523-528 [FREE Full text] [doi: 10.1016/j.brat.2011.06.012] [Medline: 21788012]
39. Santiago CD, Kaltman S, Miranda J. Poverty and mental health: how do low-income adults and children fare in psychotherapy? J Clin Psychol 2013 Feb;69(2):115-126. [doi: 10.1002/jclp.21951] [Medline: 23280880]
40. Bennett CB, Ruggero CJ, Sever AC, Yanouri L. eHealth to redress psychotherapy access barriers both new and old: a review of reviews and meta-analyses. J Psychother Integr 2020 Jun;30(2):188-207. [doi: 10.1037/int0000217]
41. Mausbach BT, Moore R, Roesch S, Cardenas V, Patterson TL. The relationship between homework compliance and therapy outcomes: an updated meta-analysis. Cognit Ther Res 2010;34(5):429-438 [FREE Full text] [doi: 10.1007/s10608-010-9297-z] [Medline: 20930925]
42. Rees CS, McEvoy P, Nathan PR. Relationship between homework completion and outcome in cognitive behaviour therapy. Cogn Behav Ther 2005;34(4):242-247. [doi: 10.1080/16506070510011548] [Medline: 16319035]
43. Eid M, Diener E. Handbook of Multimethod Measurement in Psychology. Washington, DC: American Psychological Association; 2006.
44. Baumeister H, Bauereiss N, Zarski AC, Braun L, Buntrock C, Hoherz C, et al. Clinical and cost-effectiveness of PSYCHOnlineTHERAPY: study protocol of a multicenter blended outpatient psychotherapy cluster randomized controlled




XSL•FO
RenderX

trial for patients with depressive and anxiety disorders. Front Psychiatry 2021;12:660534 [FREE Full text] [doi: 10.3389/fpsyt.2021.660534] [Medline: 34054617]

## Abbreviations

**ACT:** acceptance and commitment therapy
**BC-CBT:** blended care cognitive behavioral therapy
**CBT:** cognitive behavioral therapy
**DBT:** dialectical behavior therapy
**GAD-7:** Generalized Anxiety Disorder-7
**HIPAA:** Health Insurance Portability and Accountability Act
**iCBT:** internet-based cognitive behavioral therapy
**MAR:** missing at random
**PHQ-9:** Patient Health Questionnaire-9
**TIC:** time-invariant covariate
**TVC:** time-varying covariate

*Edited by G Eysenbach; submitted 07.10.21; peer-reviewed by A Daros; comments to author 29.10.21; accepted 21.11.21; published 17.12.21.*

*Please cite as:*
*Wu MS, Wickham RE, Chen SY, Chen C, Lungu A*
*Examining the Impact of Digital Components Across Different Phases of Treatment in a Blended Care Cognitive Behavioral Therapy Intervention for Depression and Anxiety: Pragmatic Retrospective Study*
*JMIR Form Res 2021;5(12):e33452*
*URL: https://formative.jmir.org/2021/12/e33452*
*doi:10.2196/33452*
*PMID:34927591*

©Monica S Wu, Robert E Wickham, Shih-Yin Chen, Connie Chen, Anita Lungu. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# A Patient-Reported Outcome Tool to Triage Total Hip Arthroplasty Patients to Hospital or Video Consultation: Pilot Study With Expert Panels and a Cohort of 1228 Patients

Yvette Pronk[1], MSc; Peter Pilot[2], PhD; Walter van der Weegen[3], PhD; Justus-Martijn Brinkman[4], MD, PhD; Berend Willem Schreurs[5], MD, PhD

[1]Research Department, Kliniek ViaSana, Mill, Netherlands
[2]IMUKA, Roosteren, Netherlands
[3]Department of Orthopaedic Surgery, Anna Ziekenhuis, Geldrop, Netherlands
[4]Department of Orthopaedic Surgery, Kliniek ViaSana, Mill, Netherlands
[5]Department of Orthopaedic Surgery, Radboud University Medical Center, Nijmegen, Netherlands

**Corresponding Author:**
Yvette Pronk, MSc
Research Department
Kliniek ViaSana
Hoogveldseweg 1
Mill, 5451 AA
Netherlands
Phone: 31 485476330
Email: y.pronk@viasana.nl

## Abstract

**Background:** The digital transformation in health care has been accelerated by the COVID-19 pandemic. Video consultation has become the alternative for hospital consultation. It remains unknown how to select patients suitable for video consultation.

**Objective:** This study aimed to develop a tool based on patient-reported outcomes (PROs) to triage total hip arthroplasty (THA) patients to hospital or video consultation.

**Methods:** A pilot study with expert panels and a retrospective cohort with prospectively collected data from 1228 THA patients was executed. The primary outcome was a PRO triage tool to allocate THA patients to hospital or video consultation 6 weeks postoperatively. Expert panels defined the criteria and selected the patient-reported outcome measure (PROM) questions including thresholds. Data were divided into training and test cohorts. Distribution, floor effect, correlation, responsiveness, PRO patient journey, and homogeneity of the selected questions were investigated in the training cohort. The test cohort was used to provide an unbiased evaluation of the final triage tool.

**Results:** The expert panels selected moderate or severe pain and using 2 crutches as the triage tool criteria. PROM questions included in the final triage tool were numeric rating scale (NRS) pain during activity, 3-level version of the EuroQol 5 dimensions (EQ-5D-3L) questions 1 and 4, and Oxford Hip Score (OHS) questions 6, 8, and 12. Of the training cohort, 201 (201/703, 28.6%) patients needed a hospital consultation, which was statistically equal to the 150 (150/463, 32.4%) patients in the test cohort who needed a hospital consultation ($P$=.19).

**Conclusions:** A PRO triage tool based on moderate or severe pain and using 2 crutches was developed. Around 70% of THA patients could safely have a video consultation, and 30% needed a hospital consultation 6 weeks postoperatively. This tool is promising for selecting patients for video consultation while using an existing PROM infrastructure.

***(JMIR Form Res 2021;5(12):e31232)*** doi:10.2196/31232

**KEYWORDS**

PROMs; total hip arthroplasty; triage tool; video consultation; telemedicine; digital transformation



XSL•FO
RenderX

## Introduction

The digital transformation in health care has been accelerated by the COVID-19 pandemic. Health care institutions are challenged by precautionary measures to contain COVID-19 while continuing to provide health care. Especially, managing the physical flow of patients is challenging. After the pandemic, hospitals will begin to eliminate their waiting lists while maintaining a normal patient flow, and they will be challenged again. As a solution, video consultation has become an alternative to the traditional hospital consultation. The number of hospital consultations has dropped by 30%, and the number of telemedicine visits has increased 5-fold [1].

Currently, video consultation provides health care institutions and clinicians the opportunity to increase office efficacy and cost-effectiveness in an era of decreasing reimbursements and increasing time constraints [2-4]. From the patient's perspective, it can also improve care efficacy and patient satisfaction as well as eliminating travel time and expenses [1,2]. However, not all patients might benefit from video consultation, and it is unknown how to select patients suitable for video consultation.

Orthopedic associations in many countries advise hospitals to collect patient-reported outcomes (PROs) of total hip arthroplasty (THA) using selected patient-reported outcome measures (PROMs) to evaluate health care and improve patient care [5,6]. To prevent extra burden in time and costs, it would be efficient to apply these PROs to select which patients need a hospital consultation and who can have a video consultation instead. Therefore, the aim of this study was to develop a tool based on PROs to triage THA patients to hospital or video consultation 6 weeks postoperatively. It was hypothesized that 10% of the THA patients would need a hospital consultation, as around 90% of the performed THAs result in a favorable outcome [7-9].

## Methods

### Overview

A pilot study with expert panels and a retrospective cohort with prospectively collected data from THA patients was performed. Regarding the cohort, patients were included in this study if they signed the informed consent form preoperatively to allow further scientific analysis using their anonymized data. Therefore, the institutional review board ruled that formal approval was not required for this study. There were no exclusion criteria.

### Outcomes

The primary outcome was a PRO triage tool to allocate THA patients to a hospital consultation or a video consultation for their 6-week postoperative consultation. Hospital consultation was defined as needing a physical examination or other examination, such as an X-ray, for which a patient really needed to be in the hospital. If no hospital consultation was needed, patients were allocated to a video consultation. According to the Dutch guidelines, patients should be seen 6 to 12 weeks after a THA [10], which is mostly held at 6 weeks. As it is advised to collect the first postoperative PROs at 3 months and not at 6 weeks [6], the 3-month PROs were considered the most appropriate for this study. Based on previous studies, the assumption was made that there are limited clinically relevant differences between PROs at 6 weeks and 3 months postoperatively [11,12].

### Measurements

Measurements were divided into 3 parts: (1) expert panels defined the criteria and selected the PROM questions, including the thresholds; (2) investigation of the clinimetric qualities of selected questions or triage criteria groups in the retrospective cohort with prospectively collected data; and (3) evaluation of the final triage tool.

#### *Selection by Expert Panels*

Two expert panels were created: clinical expert panel and research expert panel. The clinical expert panel consisted of 4 high-volume THA orthopedic surgeons from 2 different health care institutions. The research expert panel consisted of 3 researchers from 2 different health care institutions. As step one, the clinical expert panel defined the clinical triage criteria for the triage tool. These clinical triage criteria were based on the clinical disabilities for which patients needed to have a physical examination or other examination, such as an X-ray. As the second step, based on the clinical triage criteria, the research expert panel selected the appropriate PROM questions, including the thresholds, based on previous studies. As step three, these questions and their thresholds were presented to the clinical expert panel to discuss if these questions and/or thresholds covered the clinical triage criteria. If no threshold was reported in previous studies, the threshold was set using clinical reasoning by the clinical expert panel. These steps resulted in the PROM questions, including the thresholds, that were determined to be clinically relevant for the triage tool.

#### *Clinimetric Qualities of Selected Questions or Triage Criteria Groups*

The retrospective cohort with prospectively collected data consisted of patients who underwent surgery between January 2016 and December 2018 in a medium-sized orthopedic hospital (Kliniek ViaSana, Mill, The Netherlands). Therefore, patients were characterized by an American Society of Anesthesiologists (ASA) score of I-II and BMI ≤35. Four high-volume, experienced orthopedic surgeons performed the primary posterolateral THAs. Length of stay was generally 1 or 2 days.

The data included patient characteristics, PROM response rates, and PROs. Patient characteristics were age on the day of surgery, gender, preoperative BMI, ASA scores, and preoperative Charnley scores collected from the electronic patient records. Response rates were calculated as the number of returned questionnaires that were partially or totally completed divided by the number of THAs minus the number of THAs of patients who were deceased (returned questionnaires / [THAs - THAs of patients who were deceased]) [5]. PROs were primary digitally collected (OnlinePROMs, Rosmalen, The Netherlands). If patients were unable to handle a computer, paper questionnaires were sent. A maximum of 2 reminders to complete the PROMs were sent [13]. PROs were collected preoperatively and 3 and 12 months postoperatively according



XSL•FO
RenderX

to the advice of the Dutch Orthopedic Association. This advice included the following questionnaires: numeric rating scale (NRS) pain at rest, NRS pain during activity, 3-level version of the EuroQol 5 dimensions (EQ-5D-3L), Hip disability and Osteoarthritis Outcome Score – Physical Function Shortform (HOOS-PS), Oxford Hip Score (OHS), and an anchor question about functional improvement [6].

Pain at rest and pain during activity were both measured using an NRS question scored from 0 (no pain) to 10 (severe pain). Quality of life was assessed using the EQ-5D-3L questionnaire consisting of 2 parts: EQ visual analogue scale (EQ VAS; 0-100, with 0 as the worst imaginable health state and 100 as the best imaginable health state) and EQ-5D descriptive system existing of 5 questions about 5 dimensions. These 5 dimensions are mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, scored from 1 (no problems) to 3 (extreme problems) [14]. Furthermore, hip function was measured using the HOOS-PS questionnaire, on a scale from 0 (no difficulty) to 100 (extreme difficulty). This questionnaire consists of 5 questions scored from 0 (no difficulty) to 5 (extreme difficulty) [15,16]. Hip function and pain were assessed using the OHS questionnaire, with scores ranging from 0 (most severe symptoms) to 48 (least symptoms). This questionnaire consists of 12 questions scored from 0 (no difficulty) to 4 (extreme difficulty) [17]. Moreover, functional improvement was inquired on a 7-point Likert scale question ranging from 1 (very much deteriorated) to 7 (very much improved).

Regarding the investigation of the clinimetric qualities of selected questions or triage criteria groups by the expert panels, the cohort was divided into 2 groups: training cohort of patients who underwent surgery in 2016 or 2017 and a test cohort of patients who underwent surgery in 2018. To assess which questions were appropriate for the triage tool, the following clinimetric qualities were investigated per selected question in the training cohort: distribution, floor effect, correlation, responsiveness, and PRO patient journey. PRO patient journey was defined as a change in recovery over time. Regarding distribution, if the question did not show any distinction (median and IQR on the same level), the question was found not to be an appropriate question for the triage tool. For floor effect, if more than 15% of the patients scored the worst score [18], the question had a problem with the floor effect and was not an appropriate triage tool question. Investigating correlation, if the question was correlated ($r \geq .7$) with another selected question(s) [19], this question or one of the other(s) could be chosen instead of all the questions for the triage tool. Regarding responsiveness, if a question was not responsive ($P > .05$) [20,21], it did not distinguish well between clinical relevance and lack of clinical relevance and was not included in the tool. Furthermore, the PRO patient journeys of patients with a worse score and of patients with a better score than the threshold were investigated. If patients with a worse score on a question at 3 months scored well on that question at 12 months, this question was not included in the triage tool. To assess which questions within the selected triage criteria group (for example pain) were appropriate for the triage tool, homogeneity was investigated per triage criteria group in the training cohort. If the homogeneity increased by removing a certain question from this group (Cronbach $\alpha > .7$) [18,19], this question did not fit in this group and could be removed from the triage tool.

### *Evaluation of the Final Triage Tool*

The final triage tool was applied in the test cohort to provide an unbiased evaluation of the final tool fitted on the training dataset. Results in both cohorts were compared to investigate the hypothesis.

### Statistical Analysis

Results are reported as mean (SD), median (IQR), or n (%) based on the test performed. To investigate if there was any difference in patient characteristics, response rates, and preoperative PROs between the training and test cohorts, continuous variables were first checked for a normal distribution. Second, independent *t* tests or Mann-Whitney *U* tests for continuous variables were executed depending on the distribution of the data, and Pearson chi-square or Fisher exact tests were executed for categorical variables.

Distribution was investigated with a boxplot distribution, floor effect was determined by calculating the percentage of patients with a minimum score, and correlation was assessed with Spearman correlation analyses. Responsiveness was evaluated by performing Wilcoxon signed rank tests on the change in preoperative and 3-month scores [20,21]. The PRO patient journey of patients with a worse or better score than the threshold on a question at 3 months was evaluated by boxplot distribution at 12 months. Homogeneity was investigated with a reliability analysis, including "scale if item deleted." Before this analysis was executed, NRS pain and EQ-5D-3L questions were recoded to the same direction as the OHS questions.

Finally, the triage tool was applied for both training and test cohorts. To test the hypothesis, the numbers of hospital and video consultations for both cohorts were compared using Pearson chi-square or Fisher exact tests.

An $\alpha$ of .05 was considered statistically significant. Statistical analyses were performed using SPSS version 25.0 (IBM Corp, Armonk, NY).

## Results

### Selection by Expert Panels

"Having moderate or severe pain" and "using 2 crutches" were defined as the triage criteria by the clinical expert panel (step 1). For the criterion of "having moderate or severe pain," the research expert panel selected the following PROM questions: NRS pain at rest, NRS pain during activities, EQ-5D-3L question 4, and OHS questions 1, 8, 10, and 12. For both NRS pain questions, previous studies reported thresholds of ≤3 for no or mild pain and >3 for moderate to severe pain [22,23]. For the criterion of "using 2 crutches," EQ-5D-3L question 1 and OHS question 6 were selected (step 2). The clinical expert panel assessed the selected questions, even the NRS pain question thresholds, as appropriate. The other thresholds were discussed and defined (step 3; Table 1).



XSL•FO
RenderX

**Table 1.** Triage criteria, selected clinically relevant questions, and defined thresholds by the expert panels.

| Triage criterion and selected PROM[a] question | PROM question (score range) | Defined threshold |
|---|---|---|
| **Having moderate or severe pain** | | |
| NRS[b] pain at rest | How much pain from your hip (surgery side) did you experience at rest in the last week? (0-10) | ≥4 |
| NRS pain during activity | How much pain from your hip (surgery side) did you experience during activity in the last week? (0-10) | ≥4 |
| EQ-5D-3L[c] question 4 | Pain/discomfort (1-3) | ≥3 (extreme pain) |
| OHS[d] question 1 | During the past 4 weeks...How would you describe the pain you usually had from your hip? (0-4) | ≤1 (moderate or severe) |
| OHS question 8 | During the past 4 weeks...After a meal (sitting at a table), how painful has it been for you to stand up from a chair because of your hip? (0-4) | ≤2 (moderate, very, unbearable) |
| OHS question 10 | During the past 4 weeks...Have you had any sudden, severe pain — "shooting," "stabbing," or "spasms" — from the affected hip? (0-4) | ≤1 (most or every) |
| OHS question 12 | During the past 4 weeks...Have you been troubled by pain from your hip in bed at night? (0-4) | ≤2 (3 or 4, 5 or 6, all) |
| **Using 2 crutches** | | |
| EQ-5D-3L question 1 | Mobility (1-3) | ≥3 (confined to bed) |
| OHS question 6 | During the past 4 weeks...For how long have you been able to walk before pain from your hip becomes severe (with or without a stick)? (0-4) | ≤2 (5-15 minutes, around the house only, not at all) |

[a]PROM: patient-reported outcome measure.

[b]NRS: numeric rating scale.

[c]EQ-5D-3L: 3-level version of the EuroQol 5 dimensions.

[d]OHS: Oxford Hip Score.

## Clinimetric Qualities of Selected Questions or Triage Criteria Groups

Response rates were statistically significantly equal between both training (n=746) and test (n=482) cohorts preoperatively (745/746, 99.9% versus 482/482, 100%; *P*=.99) and at 3 months (703/746, 94.2% versus 463/482, 96.1%; *P*=.24) and 12 months (693/746, 92.9% versus 457/482, 94.8%; *P*=.29) postoperatively. The training cohort consisted of significantly fewer patients than in the test cohort with an ASA I score (399/746, 53.5% versus 287/482, 59.5%; *P*=.04), lower Charnely scores (*P*=.048), higher preoperative HOOS-PS scores (median 46.1, IQR 37.7-55.9 versus median 46.1, IQR 33.9-55.9; *P*=.01), and lower preoperative OHS scores (median 24.0, IQR 19.0-29.0 versus median 25.0, IQR 19.0-31.0; *P*=.03; Table 2). The clinical expert panel assessed these differences as not clinically relevant to correct for.



XSL•FO
RenderX

**Table 2.** Characteristics of training and test cohorts.

| Characteristics | Training cohort (n=746) | Test cohort (n=482) | *P* value |
|---|---|---|---|
| **Response rate, n (%)** | | | |
| Preoperative | 745 (99.9) | 482 (100) | .99 |
| 3 months postoperative | 703 (94.2) | 463 (96.1) | .24 |
| 12 months postoperative | 693 (92.9) | 457 (94.8) | .29 |
| **Patient characteristics** | | | |
| Age (years), median (IQR) | 66.5 (61.0-72.0) | 66.0 (60.0-72.0) | .73 |
| Gender (male), n (%) | 295 (39.5) | 206 (42.7) | .27 |
| BMI (kg/m$^2$), median (IQR) | 26.1 (24.1-28.4) | 26.3 (24.1-28.5) | .48 |
| ASA[a] score – I, n (%) | 399 (53.5) | 287 (59.5) | .04 |
| **Charnley score:** | | | .048 |
| One hip affected by OA[b], n (%) | 163 (21.8) | 108 (22.4) | |
| Both hips affected by OA, n (%) | 309 (41.4) | 196 (40.7) | |
| Contralateral hip OA, n (%) | 163 (21.8) | 82 (17.0) | |
| Multiple joints affected by OA, n (%) | 111 (14.9) | 96 (19.9) | |
| **Preoperative PROs[c], median (IQR)** | | | |
| NRS[d] pain at rest score | 6.0 (4.0-7.0) | 6.0 (3.8-7.0) | .18 |
| NRS pain during activity score | 8.0 (7.0-9.0) | 8.0 (7.0-8.0) | .13 |
| HOOS-PS[e] score | 46.1 (37.7-55.9) | 46.1 (33.9-55.9) | .01 |
| EQ-5D descriptive system[f] | 0.693 (0.298-0.775) | 0.693 (0.569-0.775) | .16 |
| EQ VAS[g] | 76.0 (63.3-89.8) | 77.0 (60.0-86.3) | .82 |
| OHS[h] | 24.0 (19.0-29.0) | 25.0 (19.0-31.0) | .03 |

[a]ASA: American Society of Anesthesiologists.
[b]OA: osteoarthritis.
[c]PROs: patient-reported outcomes.
[d]NRS: numeric rating scale.
[e]HOOS-PS: Hip disability and Osteoarthritis Outcome Score — Physical Function Shortform.
[f]EQ-5D descriptive system: EuroQol 5 dimensions descriptive system.
[g]EQ VAS: EuroQol visual analogue scale.
[h]OHS: Oxford Hip Score.

Regarding the questions or triage criteria groups selected by the expert panels (Table 1), OHS question 10 showed no distribution. For floor effect, <15% of patients scored the minimum score on all questions separately. All questions were significantly correlated with each other ($P$<.001). Regarding correlations ≥0.7, NRS pain during activity correlated with NRS pain at rest ($r$=0.659, $P$<.001) and OHS question 1 ($r$=–0.676, $P$<.001; Table 3). Furthermore, all questions were shown to be responsive ($P$<.001; Table 4). Regarding the PRO patient journey, patients with a worse score than the threshold also reported worse scores at 12 months than patients with a better score than the threshold. Only one patient with a better score than the threshold on EQ-5D-3L question 1 at 3 months had a 12-month score (Table 5). The other questions included ≥11 patients below or above the threshold. Regarding homogeneity, a Cronbach α of 0.818 was found for the triage criteria group "pain." When one of the questions in this group was removed, the Cronbach α was maintained at above 0.7. The triage criteria group "crutches" scored a Cronbach α of 0.628. As there were 2 questions in this group, none of them could be removed to investigate the Cronbach α.



XSL•FO
RenderX

**Table 3.** Distribution, floor effect, and correlation per selected patient-reported outcome measure (PROM) question.

| PROM[a] question | Distribution, median (IQR) | Floor effect, n (%) | Correlations | | |
|---|---|---|---|---|---|
| | | | Correlations (*r*)[b] | Correlated question | *P* value |
| NRS[c] pain at rest | 0 (0-1) | 3[d] (0.4) | 0.659 | NRS pain during activity | <.001 |
| NRS pain during activity | 2 (0-3) | 4[d] (0.6) | 0.659; –0.676 | NRS pain at rest; OHS[e] question 1 | Both <.001 |
| EQ-5D-3L[f] question 4 | 1 (1-2) | 13[g] (1.9) | None | N/A[h] | N/A |
| OHS question 1 | 3 (3-4) | 6[g] (0.9) | –0.675 | NRS pain during activity | <.001 |
| OHS question 8 | 3 (3-4) | 0[g] (0.0) | None | N/A | N/A |
| OHS question 10 | 4 (4-4) | 2[i] (0.3) | None | N/A | N/A |
| OHS question 12 | 4 (3-4) | 32[i] (4.6) | None | N/A | N/A |
| EQ-5D-3L question 1 | 1 (1-2) | 2[i] (0.3) | None | N/A | N/A |
| OHS question 6 | 4 (3-4) | 3[j] (0.4) | None | N/A | N/A |

[a]PROM: patient-reported outcome measure.
[b]Statistically significant correlations >0.6 or <–0.6 are presented.
[c]NRS: numeric rating scale.
[d]n=703.
[e]OHS: Oxford Hip Score.
[f]EQ-5D-3L: 3-level version of the EuroQol 5 dimensions.
[g]n=693.
[h]N/A: not applicable.
[i]n=694.
[j]n=690.

**Table 4.** Responsiveness for each patient-reported outcome measure (PROM) question.

| PROM question | Preoperative, median (IQR) | 3 months postoperative, median (IQR) | *P* value |
|---|---|---|---|
| NRS[a] pain at rest | 6 (4-7) | 0 (0-1) | <.001 |
| NRS pain during activity | 8 (7-9) | 2 (0-3) | <.001 |
| EQ-5D-3L[b] question 4 | 2 (2-3) | 1 (1-2) | <.001 |
| OHS[c] question 1 | 1 (0-1) | 3 (3-4) | <.001 |
| OHS question 8 | 2 (2-3) | 3 (3-4) | <.001 |
| OHS question 10 | 2 (1-3) | 4 (4-4) | <.001 |
| OHS question 12 | 2 (0-3) | 4 (3-4) | <.001 |
| EQ-5D-3L question 1 | 2 (2-2) | 1 (1-2) | <.001 |
| OHS question 6 | 2 (2-3) | 4 (3-4) | <.001 |

[a]NRS: numeric rating scale.
[b]EQ-5D-3L: 3-level version of the EuroQol 5 dimensions.
[c]OHS: Oxford Hip Score.



XSL•FO
RenderX

**Table 5.** Patient journey per patient-reported outcome measure (PROM) question.

| PROM question | Defined threshold | 12-month score of patients with a score below the threshold at 3 months, median (IQR) | 12-month score of patients with a score above the threshold at 3 months, median (IQR) |
|---|---|---|---|
| NRS[a] pain at rest | ≥4 | 0 (0-1) | 2 (0-5) |
| NRS pain during activity | ≥4 | 0 (0-1) | 2 (0-4) |
| EQ-5D-3L[b] question 4 | ≥3 | 1 (1-2) | 2 (1.5-2) |
| OHS[c] question 1 | ≤1 | 3 (2-4) | 4 (3-4) |
| OHS question 8 | ≤2 | 3 (3-4) | 4 (4-4) |
| OHS question 10 | ≤1 | 4 (2.5-3) | 4 (4-4) |
| OHS question 12 | ≤2 | 3 (1.5-4) | 4 (4-4) |
| EQ-5D-3L question 1 | ≥3 | 1 (1-2) | 1 (1-1)[d] |
| OHS question 6 | ≤2 | 3 (2-4) | 4 (4-4) |

[a]NRS: numeric rating scale.

[b]EQ-5D-3L: 3-level version of the EuroQol 5 dimensions.

[c]OHS: Oxford Hip Score.

[d]n=1.

Based on the clinimetric qualities of selected questions or triage criteria groups, NRS pain at rest, OHS question 1, and OHS question 10 were removed from the triage tool. The final triage tool consisted of NRS pain during activity; EQ-5D-3L questions 1 and 4; and OHS questions 6, 8, and 12.

### Evaluation of the Final Triage Tool

The final triage tool resulted in 201 (201/703, 28.6%) patients in the training cohort needing a hospital consultation, which was statistically equal to the 150 (150/463, 32.4%) patients in the test cohort who needed a hospital consultation ($P$=.19).

## Discussion

This study aimed to develop a tool based on PROs collected using an existing PROM infrastructure to triage THA patients to hospital or video consultation 6 weeks postoperatively. As the main finding, a triage tool based on PROM questions measuring moderate or severe pain and whether the patient used 2 crutches was developed. The included questions were NRS pain during activity; EQ-5D-3L questions 1 and 4; and OHS questions 6, 8, and 12. Applying the final triage tool in both the training and test cohorts resulted in the same outcome: Around 70% of the patients could safely have a video consultation, and 30% needed to have a hospital consultation 6 weeks postoperatively. Therefore, this PRO triage tool is a promising instrument to select patients for video consultation while using an existing PROM infrastructure. The next step is to further investigate this triage tool in daily practice.

This study showed that 70% of the hospital consultations for THA patients 6 weeks postoperatively could safely be done by video. It was hypothesized that 10% of the THA patients would need a hospital consultation. First, the result that 30% of patients needed a hospital consultation could be explained by the focus of the clinical expert panel. As the experts' beginning point was seeing all patients during a hospital consultation (100%), by developing the triage tool, they desired to see all patients who potentially needed a physical examination or other examination, such as an X-ray, during a hospital consultation. Furthermore, they desired to prevent obtaining more consultations by needing to schedule a hospital consultation after a video consultation. Both implicitly resulted in more liberal criteria for a hospital consultation leading to more patients triaged to a hospital consultation. Second, it could be that specific questions are missing from the triage tool. It was hypothesized that, after an investigation of the triage tool in daily practice, the criteria for the triage tool could be improved, achieving the right health care for each patient and a further reduction in hospital consultations. It would be interesting to investigate how many additional hospital consultations would be needed if the tool triages to video consultation and which PROs are different for patients with an additional hospital consultation.

It is essential to understand that the PRO triage tool is not a tool on its own, but it is the first step in the selection of patients who need a hospital consultation and those who can have a video consultation instead. PROs and clinical judgment produce complementary data and when combined, provide a more accurate description of the patients' symptoms [24]. Therefore, using the current PRO triage tool, clinicians should have the ability to change the outcome of the tool. To further develop the triage tool, it would be interesting to investigate how many times clinicians decide to change the outcome and which PROs are different for the patients whose clinicians decide to change the outcome. Furthermore, it would be interesting to take the patient's preference into account.

Previous studies reported that patients rate their video consultations as excellent or very good (92%-95%) [25,26]. The patient no-show rate has been reported at 2.8%, and their mean estimated saved travel time is 30 minutes [25]. Furthermore, 82% would recommend video consultation to family and friends [26]. Almost all clinicians rate their video consultation experience as very good or excellent (92%). They are comfortable with executing this type of consultation after



XSL•FO
RenderX

1 to 4 sessions (69%) [25]. Therefore, video consultation is a serious alternative for hospital consultation. Numbers could be improved when appropriate patients for video consultation are selected, which makes the developed PRO triage tool a promising instrument.

As a first strength of this study, to the authors' knowledge, this is the first study in which a tool to triage patients to hospital or video consultation was developed. Second, high response rates preoperatively and even postoperatively (above 90%) were achieved, resulting in a representative cohort to execute this study and to generalize the results to the total THA population. A third strength is the application of the training and test cohorts to provide an unbiased evaluation of the final tool.

As a limitation of this study, the triage tool was not investigated in a prospective cohort, and aspects of reliability, validity, sensitivity, and specificity of the triage tool were not investigated yet. Furthermore, 3-month PROs were used instead of 6-week PROs, as, although based on previous studies, the assumption was made that there is limited clinically relevant difference between PROs at 6 weeks and at 3 months postoperatively [11,12]. Future research should be executed in a prospective cohort, and aspects of reliability, validity, sensitivity, and specificity need to be investigated to further develop the THA PRO triage tool. The triage tool could be improved by investigating which PROs are different for patients with additional hospital consultations after being triaged to video consultation or for patients whose clinician decided to change the outcome of the triage tool. After improving the triage tool, it would be interesting to investigate if and which of the patients triaged to video consultation may not require a consultation at all.

In conclusion, a THA PRO triage tool based on moderate or severe pain and using 2 crutches was developed. Around 70% of THA patients could safely have a video consultation, and 30% of patients needed a hospital consultation 6 weeks postoperatively. This tool is promising for selecting patients for video consultation while using an existing PROM infrastructure.

## Acknowledgments

We would like to thank Klaartje van Diepen-Pijnappels and Maud Peters for their consistent data collection and kindness helping all patients who had questions, all patients for completing their PROMs, and the orthopedic surgeons for participating in the clinical expert panel.

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

## Authors' Contributions

YP was responsible for designing the study, data collection, statistical analysis, and study coordination. All authors were involved in drafting and revising the manuscript and approved the final version of the manuscript.

## Conflicts of Interest

PP was a paid employee of ZimmerBiomet, and WW is a consultant for ZimmerBiomet, which have nonfinancial associations that may be relevant to the submitted manuscript. BWS receives a grant for providing an educational Stryker course on the Exeter prosthesis and is a board member of the European Hip Society, which has nonfinancial associations that may be relevant to the submitted manuscript. The other authors declare that they have no conflict of interest.

## References

1. Merkes K, Hagenaars N. Corona: katalysator of struikelblok voor groenere ziekenhuiszorg? Gupta Strategists. 2020 May 28. URL: https://gupta-strategists.nl/studies/corona-katalysator-of-struikelblok-voor-groenere-ziekenhuiszorg [accessed 2021-11-13]
2. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
3. O'Connor CM, Anoushiravani AA, DiCaprio MR, Healy WL, Iorio R. Economic recovery after the COVID-19 pandemic: resuming elective orthopedic surgery and total joint arthroplasty. J Arthroplasty 2020 Jul;35(7S):S32-S36 [FREE Full text] [doi: 10.1016/j.arth.2020.04.038] [Medline: 32345566]
4. Kane LT, Thakar O, Jamgochian G, Lazarus MD, Abboud JA, Namdari S, et al. The role of telehealth as a platform for postoperative visits following rotator cuff repair: a prospective, randomized controlled trial. J Shoulder Elbow Surg 2020 Apr;29(4):775-783. [doi: 10.1016/j.jse.2019.12.004] [Medline: 32197766]
5. Rolfson O, Bohm E, Franklin P, Lyman S, Denissen G, Dawson J, Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries. Report of the Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part II. Recommendations for selection, administration, and analysis. Acta Orthop 2016 Jul;87 Suppl 1:9-23 [FREE Full text] [doi: 10.1080/17453674.2016.1181816] [Medline: 27228230]
6. PROMs. Nederlandse Orthopaedische Vereniging. URL: https://www.lroi.nl/invoerders/registreren/proms [accessed 2018-08-01]



XSL•FO
RenderX

7. Palazzo C, Jourdan C, Descamps S, Nizard R, Hamadouche M, Anract P, et al. Determinants of satisfaction 1 year after total hip arthroplasty: the role of expectations fulfilment. BMC Musculoskelet Disord 2014 Feb 24;15:53 [FREE Full text] [doi: 10.1186/1471-2474-15-53] [Medline: 24564856]
8. Anakwe RE, Jenkins PJ, Moran M. Predicting dissatisfaction after total hip arthroplasty: a study of 850 patients. J Arthroplasty 2011 Feb;26(2):209-213. [doi: 10.1016/j.arth.2010.03.013] [Medline: 20462736]
9. Hesseling B, Mathijssen NMC, van Steenbergen LN, Melles M, Vehmeijer SBW, Porsius JT. Fast starters, slow starters, and late dippers: trajectories of patient-reported outcomes after total hip arthroplasty: results from a Dutch nationwide database. J Bone Joint Surg Am 2019 Dec 18;101(24):2175-2186. [doi: 10.2106/JBJS.19.00234] [Medline: 31609887]
10. Routinematige follow-up bij THP - Richtlijnendatabase. Federatie Medisch Specialisten. 2019 Dec 2. URL: https://richtlijnendatabase.nl/richtlijn/totale_heupprothese_thp/postoperatieve_zorg_bij_thp/routinematige_follow-up_bij_thp.html [accessed 2021-11-13]
11. Klapwijk LCM, Mathijssen NMC, Van Egmond JC, Verbeek BM, Vehmeijer SBW. The first 6 weeks of recovery after primary total hip arthroplasty with fast track. Acta Orthop 2017 Apr;88(2):140-144 [FREE Full text] [doi: 10.1080/17453674.2016.1274865] [Medline: 28079428]
12. Hartog YMD, Mathijssen NMC, Vehmeijer SBW. Total hip arthroplasty in an outpatient setting in 27 selected patients. Acta Orthop 2015;86(6):667-670 [FREE Full text] [doi: 10.3109/17453674.2015.1066211] [Medline: 26139431]
13. Pronk Y, Pilot P, Brinkman JM, van Heerwaarden RJ, van der Weegen W. Response rate and costs for automated patient-reported outcomes collection alone compared to combined automated and manual collection. J Patient Rep Outcomes 2019 Jun 03;3(1):31 [FREE Full text] [doi: 10.1186/s41687-019-0121-6] [Medline: 31155689]
14. EuroQol Group. EuroQol–a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
15. Davis AM, Perruccio AV, Canizares M, Hawker GA, Roos EM, Maillefert J, et al. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. Osteoarthritis Cartilage 2009 Jul;17(7):843-847 [FREE Full text] [doi: 10.1016/j.joca.2009.01.005] [Medline: 19215728]
16. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008 May;16(5):551-559 [FREE Full text] [doi: 10.1016/j.joca.2007.12.016] [Medline: 18296074]
17. Gosens T, Hoefnagels NHM, de Vet RCW, Dhert WJA, van Langelaan EJ, Bulstra SK, et al. The "Oxford Heup Score": the translation and validation of a questionnaire into Dutch to evaluate the results of total hip arthroplasty. Acta Orthop 2005 Apr;76(2):204-211. [doi: 10.1080/00016470510030580] [Medline: 16097545]
18. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
19. Streiner DL, Norman GR, Cairney J. Reliability. In: Streiner DL, Norman GR, Cairney J, editors. Health Measurement Scales: A practical guide to their development and use. Oxford, England: Oxford University Press; 2019. [doi: 10.1093/med/9780199685219.003.0008]
20. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 2003 Jun;12(4):349-362. [doi: 10.1023/a:1023499322593] [Medline: 12797708]
21. Moayyedi P, Duffett S, Braunholtz D, Mason S, Richards ID, Dowell AC, et al. The Leeds Dyspepsia Questionnaire: a valid tool for measuring the presence and severity of dyspepsia. Aliment Pharmacol Ther 1998 Dec;12(12):1257-1262 [FREE Full text] [doi: 10.1046/j.1365-2036.1998.00404.x] [Medline: 9882035]
22. Pinto PR, McIntyre T, Ferrero R, Almeida A, Araújo-Soares V. Risk factors for moderate and severe persistent pain in patients undergoing total knee and hip arthroplasty: a prospective predictive study. PLoS One 2013;8(9):e73917 [FREE Full text] [doi: 10.1371/journal.pone.0073917] [Medline: 24058502]
23. Breivik H, Borchgrevink PC, Allen SM, Rosseland LA, Romundstad L, Hals EKB, et al. Assessment of pain. Br J Anaesth 2008 Jul;101(1):17-24 [FREE Full text] [doi: 10.1093/bja/aen103] [Medline: 18487245]
24. Basch E, Bennett A, Pietanza MC. Use of patient-reported outcomes to improve the predictive accuracy of clinician-reported adverse events. J Natl Cancer Inst 2011 Dec 21;103(24):1808-1810. [doi: 10.1093/jnci/djr493] [Medline: 22157639]
25. Tenforde AS, Iaccarino MA, Borgstrom H, Hefner JE, Silver J, Ahmed M, et al. Telemedicine during COVID-19 for outpatient sports and musculoskeletal medicine physicians. PM R 2020 Sep;12(9):926-932 [FREE Full text] [doi: 10.1002/pmrj.12422] [Medline: 32424977]
26. Donelan K, Barreto EA, Sossong S, Michael C, Estrada JJ, Cohen AB, et al. Patient and clinician experiences with telehealth for patient follow-up care. Am J Manag Care 2019 Jan;25(1):40-44 [FREE Full text] [Medline: 30667610]

## Abbreviations

**ASA:** American Society of Anesthesiologists
**EQ VAS:** EuroQol visual analogue scale



XSL•FO
RenderX

**EQ-5D descriptive system:** EuroQol 5 dimensions descriptive system
**EQ-5D-3L:** 3-level version of the EuroQol 5 dimensions
**HOOS-PS:** Hip disability and Osteoarthritis Outcome Score – Physical Function Shortform
**NRS:** numeric rating scale
**OHS:** Oxford Hip Score
**PROM(s):** patient-reported outcome measure(s)
**PRO(s):** patient-reported outcome(s)
**THA:** total hip arthroplasty

*Edited by G Eysenbach; submitted 14.06.21; peer-reviewed by DP Shah, B McKinstry; comments to author 07.09.21; revised version received 04.10.21; accepted 04.10.21; published 20.12.21.*

*Please cite as:*
*Pronk Y, Pilot P, van der Weegen W, Brinkman JM, Schreurs BW*
*A Patient-Reported Outcome Tool to Triage Total Hip Arthroplasty Patients to Hospital or Video Consultation: Pilot Study With Expert Panels and a Cohort of 1228 Patients*
*JMIR Form Res 2021;5(12):e31232*
*URL: https://formative.jmir.org/2021/12/e31232*
*doi:10.2196/31232*
*PMID:34931989*

©Yvette Pronk, Peter Pilot, Walter van der Weegen, Justus-Martijn Brinkman, Berend Willem Schreurs. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# The Successful Synchronized Orchestration of an Investigator-Initiated Multicenter Trial Using a Clinical Trial Management System and Team Approach: Design and Utility Study

Dinesh Pal Mudaranthakam[1], MBA; Alexandra Brown[1], MS; Elizabeth Kerling[1], MS; Susan E Carlson[1], PhD; Christina J Valentine[2], MD; Byron Gajewski[1], PhD

[1]University of Kansas Medical Center, Kansas City, KS, United States
[2]University of Cincinnati, Cincinnati, OH, United States

**Corresponding Author:**
Dinesh Pal Mudaranthakam, MBA
University of Kansas Medical Center
3901 Rainbow Blvd
Kansas City, KS, 66160
United States
Phone: 1 9139456922
Email: dmudaranthakam@kumc.edu

## Abstract

**Background:** As the cost of clinical trials continues to rise, novel approaches are required to ensure ethical allocation of resources. Multisite trials have been increasingly utilized in phase 1 trials for rare diseases and in phase 2 and 3 trials to meet accrual needs. The benefits of multisite trials include easier patient recruitment, expanded generalizability, and more robust statistical analyses. However, there are several problems more likely to arise in multisite trials, including accrual inequality, protocol nonadherence, data entry mistakes, and data integration difficulties.

**Objective:** The Biostatistics & Data Science department at the University of Kansas Medical Center developed a clinical trial management system (comprehensive research information system [CRIS]) specifically designed to streamline multisite clinical trial management.

**Methods:** A National Institute of Child Health and Human Development–funded phase 3 trial, the ADORE (assessment of docosahexaenoic acid [DHA] on reducing early preterm birth) trial fully utilized CRIS to provide automated accrual reports, centralize data capture, automate trial completion reports, and streamline data harmonization.

**Results:** Using the ADORE trial as an example, we describe the utility of CRIS in database design, regulatory compliance, training standardization, study management, and automated reporting. Our goal is to continue to build a CRIS through use in subsequent multisite trials. Reports generated to suit the needs of future studies will be available as templates.

**Conclusions:** The implementation of similar tools and systems could provide significant cost-saving and operational benefit to multisite trials.

**Trial Registration:** ClinicalTrials.gov NCT02626299; https://tinyurl.com/j6erphcj

***(JMIR Form Res 2021;5(12):e30368)*** doi:10.2196/30368

**KEYWORDS**

data management; data quality; metrics; trial execution; clinical trials; cost; accrual; accrual inequality; rare diseases; healthcare; health care; health operations



XSL•FO
RenderX

## Introduction

### Background

The evaluation of new treatment modalities through randomized controlled trials (RCTs) is the gold standard for advancing medical research; however, RCTs frequently fail to meet recruitment goals [1]. Multisite trials are frequently utilized to increase the availability of patients who meet inclusion criteria in phase 3 trials and increasingly in phase 1 and phase 2 trials as well. The benefits of performing multisite trials include more robust statistical analyses, reduced bias, expanded generalizability, and shorter recruitment periods [2,3]. However, multisite trials are far more difficult to perform effectively than single-site trials.

The primary research team and site leads have an obligation to ensure the trial is conducted in accordance with the International Council for Harmonization Guideline for Good Clinical Practice (ICH GCP) [4]. Standardized training and a central data management system are necessary to overcome the variability in clinical trial experience among study sites, which commonly have different recruiting practices, data entry processes, regulatory interpretation, and clinical trial experience [5,6]. Without standardization, these differences can threaten data harmonization, quality assurance, statistical analysis, data management, regulatory compliance, and recruitment [7]. A central data management system combined with standardized data entry protocols and customized reporting that includes ongoing accrual tracking [8,9] can facilitate data harmonization, quality assurance, data cleaning, and data analysis.

We detail here a clinical trial management system (CTMS), powered by WCG (Western Institutional Review Board–Copernicus Group, Inc) Velos and customized by the University of Kansas Medical Center (KUMC) Biostatistics & Data Science department to address the challenges of multisite trial management. The CTMS (comprehensive research information system [CRIS]) was customized to maximize team collaboration and produce effective workflow procedures for a phase 3, multisite RCT (the ADORE trial) in partnership with the Maternal and Child Health team at the Department of Dietetics and Nutrition at KUMC.

### The RCT in Brief

The ADORE (assessment of docosahexaenoic acid [DHA] on reducing early preterm birth) trial was a phase 3, investigator-initiated, adaptive-design, double-blind randomized, superiority trial designed to determine the potential of high dose DHA to reduce the incidence of early preterm birth (<34 weeks) (R01HD083292; HD NCT02626299). Pregnant women were assigned to either a standard prenatal DHA supplement of 200 mg per day or 1000 mg per day before 20 weeks gestation [10]. The study sought to enroll between 900 and 1200 pregnant women over a 4-year period, which lent itself naturally to a multisite configuration.

Three academic medical centers enrolled a total of 1100 participants in ADORE: the KUMC (n=489), University of Cincinnati (UC; n=252), and the Ohio State University (OSU; n=359). During the study, 128 employees were involved across the 3 trial teams.

While this paper focuses on a specific CTMS and RCT, the overall purpose is to characterize the strengths and weaknesses of a centralized data management system and its integration into multisite trials. The broad topics of database design, regulatory compliance, CTMS training and access, study management, and data management are emphasized throughout manuscript.

## Methods

### Methods Overview

The CTMS deployed at the KUMC was powered by WCG Velos and customized by the Biostatistics & Data Science department to include features related to randomization, automated reports, dashboards, etc, and titled the CRIS. Features such as electronic data capture, data monitoring, and data validation are common in clinical trial management systems. The value of the CRIS system was to standardize these aspects of clinical trial management systems for multiple sites, as well as introduce additional utility uniquely suited to multisite clinical trials.

### Development of an Electronic Case Report Form

Database design was an integral piece of the protocol development phase of study design. It was imperative to understand the capabilities of CRIS in collaboration with key personnel well before recruitment began. Critical factors to address included the number of study sites, sample size, number of treatment arms, and recruitment protocol. Senior members of the primary KUMC site team including principal investigators, the project director, biostatistician, and CTMS director of research information technology met regularly in person during the design phase. Primary, secondary, and tertiary data points were examined carefully, including their variable type and validation. Study personnel at satellite recruitment sites in Ohio additionally vetted data fields by verifying that the data elements that were part of the electronic case report form (eCRF) were aligned with the aims of the protocol. This ensured that these fields were able to capture the correct values for the study and allowed data capture to occur as categorical values rather than open text fields. Adverse event (AE)–reporting and participant-focused activity-tracking including standardized recruitment, randomization, blinding, treatment follow-up, and specimen collection forms were also considered in the CRIS database design prior to study start. An eCRF was an outcome of the design phase.

### Achieving Regulatory Compliance

In addition to the standard elements of ICH GCPs, eCRFs were required to be compliant with 21 CFR Part 11 *Electronic Records; Electronic Signature* to ensure security and accountability of study data [11,12]. Requirements of this compliance included controls for open systems, signature manifestations, record-linking, and controls for identification codes and passwords. Open system controls included continuous system validation, limitation of system access, computer-generated audit trails, operational system checks, accountability in record changes, document encryption, and



XSL•FO
RenderX

several other controls. Signature controls ensured that any changes to records did not obscure previous information, that changes were associated the study member who made them, and that the date and time of the change was recorded. These controls, and several others, were integrated into CRIS development to ensure full CFR Part 11 compliance during electronic data capture.

## Study Personnel Training

The training of study personnel at all participating sites followed design of the eCRF. Both live and recorded sessions reviewed the protocol, the Manual of Operating Procedures, and the CRIS system and associated eCRF. Access to the system was provided only after users had completed both institutionally required ethics and study-specific training. Access within the system was role-based; that is, principal investigator, co-investigator, study coordinator, monitor, pharmacists, etc, to ensure study team members were only able to access data within CRIS relevant to their role on the trial. While the training itself was conducted outside of the CRIS through the use of prerecorded training videos and manual materials, study staff were tested on the material, and their testing results and training certification were documented within the CRIS system and was a prerequisite for system access.

The study operational director and primary principal investigator visited the participating sites before enrollment began, and the operational director made additional visits during the study to provide additional training and tips and to audit study documentation and workflow.

## Data Management

Data entry was performed at each site via the web-based browser and managed at KUMC through the Biostatistics & Data Science department. This was integrated to allow personnel at all sites to conserve time and labor typically allocated to data management. The benefit of eCRFs to data integrity, time management, and data analysis have been described [13,14].

The customized CRIS eCRFs were the primary repository for participant's historical information obtained from either health records or participant self-reports. Individual forms used during the study included forms for patient characteristics, laboratory samples, primary outcome variables, and patient participation. Patient characteristics forms included assessments of health history, dietary supplement intake, maternal physical exam, and medical record review. These forms were completed during the enrollment visit after participants gave consent for the study.

Several eCRFs were utilized during the study including forms to record samples obtained (maternal blood draws, pregnancy labs, and urine collection) and to track results of pregnancy outcomes and estimated date of birth by study site. Lastly, forms detailing patient participation were integrated in the system including the signed informed consent, study coordinator sign off, forms for the delivery and tracking of treatment, withdrawal forms, and forms documenting adverse events. All forms could be viewed, selected, and copied in list form (Multimedia Appendix 1). Paper copies of source documents to verify data entry were retained by each study site and scanned copies submitted to CRIS for cross-validation and regulatory compliance.

## Randomization

The initial randomization table was generated by the statistical team in accordance with the study protocol guidelines. The data management team uploaded the randomization table under the CRIS. Every time a new participant was screened and deemed eligible, the participant at the time of enrollment was then randomized to a treatment arm on the basis of the stratification variables that were built under the randomization form. Once the randomization form was completed by the study coordinator the participant automatically would get randomized and the arm to which the participant was randomized would be displayed on the participants profile.

## Assuring Data Integrity

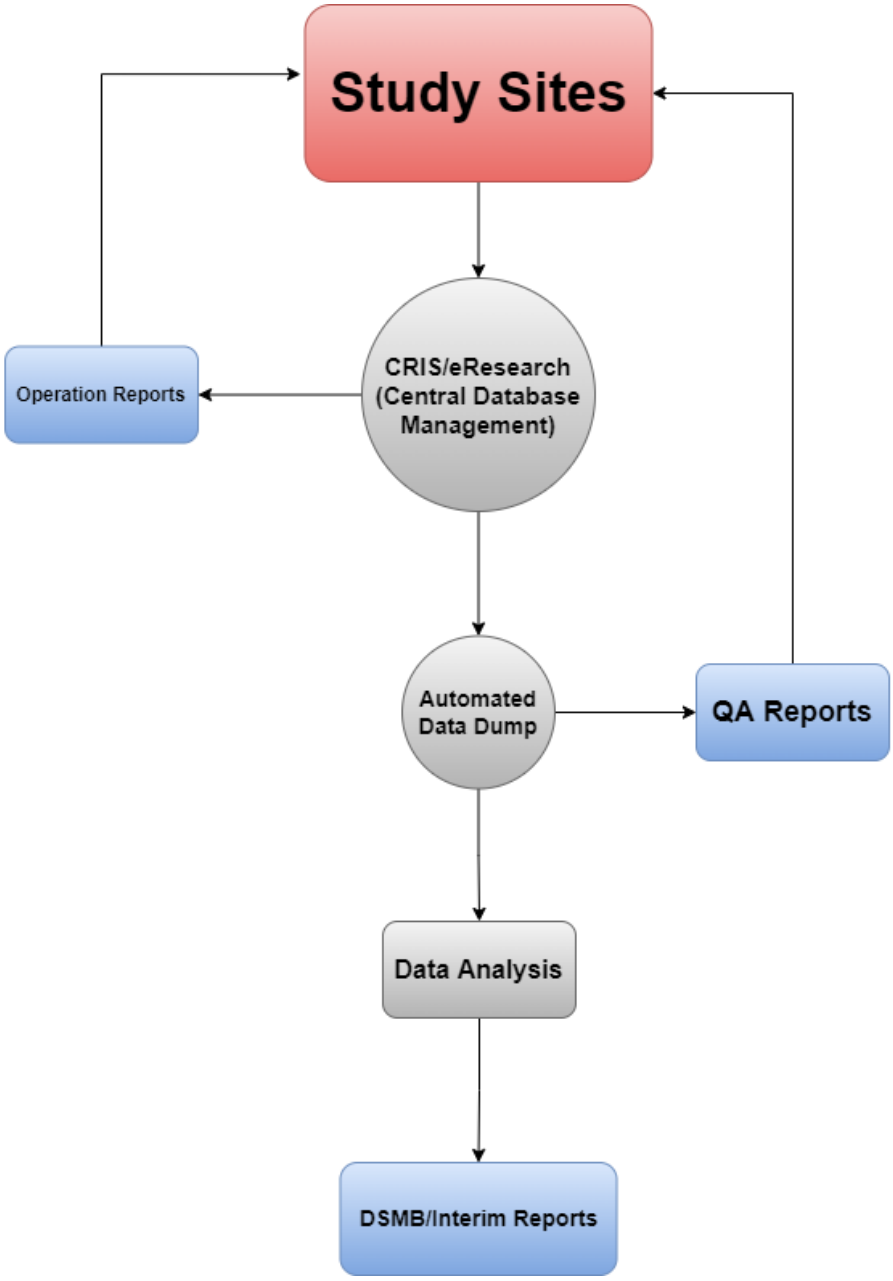
The accuracy of data entry was confirmed using a 2-pass approach whenever possible. Data entered into the eCRF and the source documentation uploaded as a secure file at satellite sites were accessible to team members at the main trial site who reviewed the eCRF for accuracy in comparison with the source document. In compliance with ICH GCP (version 4.9.1), study data were fully accessible to principal investigators responsible for evaluation of accuracy, completeness, and legibility of entered data. The ability to access the data was included to allow study coordinators to keep in contact with study participants at each site and quickly resolve any questions or concerns the study team or participant might have had.

Data harmonization is a significant hurdle for multisite studies but was aided here by having a central database and standardized data entry forms, Data harmonization occurred centrally at the KUMC utilizing CRIS and through the KUMC Biostatistics & Data Science team. Following data entry and validation at individual sites, the study results for each site were validified by regular performance of edit, logic, and range checks by the trial analyst. Queries were then sent to the clinical team at each site in the form of weekly automatically generated emails to resolve any discrepancies. All queries were resolved before the trial analyst created data sets for the interim and final analyses. The finalized study binder was produced in collaboration with the KUMC Director of Research Information Technology and contained copies of the annotated project case report forms, final data dictionary, and copies of electronic data files. This process can be visualized in Figure 1.



XSL•FO
RenderX

**Figure 1.** Architecture diagram.




All the raw participant information is captured by the site coordinators in real time with the participant in the room. Weekly reports are then generated and automatically pushed to the study team members to help keep them abreast of the study progression. Additionally, automated quality assurance reports help the quality analysis team to verify and address data inconsistences. Finally, the automated data dump is also utilized by the statistical team to perform data analysis, which is then utilized for Data Safety and Monitoring Board (DSMB).

## Study Management

The primary aim of CRIS in study management was to increase procedural uniformity across multiple stages of participant engagement. Research teams at all sites reviewed eligibility criteria, including age, gestation length, and multiple gestation, in CRIS prior to enrollment. After consent, enrollment and randomization was captured using CRIS. Each site was expected to enroll 2-3 participants weekly, and integration and management of accrual data occurred centrally at the KUMC. Each of the 3 sites were given a separate randomization code, and participants at each site were randomized using a Bayesian Adaptive Design detailed by Brown et al [15]. Subsequent adaptations and adapted randomization schedules were appended to each site within CRIS. The ability of CRIS to ensure accrual equality across study sites was assessed using a Gini coefficient. Haidich et al [16] proposed that a Gini coefficient for accrual distribution in multisite trials could provide a standardized approach to assessing accrual disparities. They identified a Gini coefficient of less than 0.2 as suggesting low accrual inequality and calculated a mean Gini coefficient of 0.33 among multisite trials.

## Ethics Approval

The KUMC granted approval under a central institutional review board with reliance by the other institutions (STUDY00003455).



XSL•FO
RenderX

The trial was registered on ClinicalTrials.gov (NCT02626299) on December 8, 2015.

## Results

### Results Overview

Features such as electronic data capture, data monitoring, and data validation are common in clinical trial management systems. An additional benefit of implementing the CRIS in this trial came from automated reports for trial accrual, study protocol adherence, and data validation tailored to the challenges of multisite trials. Accrual reports were developed to ensure consistent accrual, accrual equality between sites, and to provide accrual predictions. Study protocol reports were developed to ensure protocol adherence and prepare study teams for upcoming responsibilities. Data validation reports were generated to maintain data integrity and assist in data harmonization. These reports allowed the CRIS to provide additional support to the trial team.

### Trial Accrual Reports

Accrual equality *across* study sites is important in multisite trials as it helps capture a much broader picture, allowing study teams to generalize easily. The calculated Gini coefficient for the ADORE trial was 0.14, indicating a low accrual inequality among sites with 359 patients enrolled from the OSU sites, 489 patients from the KUMC sites, and 252 patients from the UC site. The presence of low accrual inequality in this study suggests that the generalizability of study results was maintained. Enrollment at KUMC began 3 months prior to OSU and 5 months before UC.

Successful accrual in the ADORE trial relied heavily on the weekly accrual and delivery reports generated by the CRIS. The reports began with a summary of current enrollment and delivery figures, including predictions for accrual goal achievement and 95% CIs. Participant status was outlined, including participants receiving treatment as well as those who had completed the study or discontinued treatment early. Accrual by site was available to view in the current month (Table 1), past 30 days, and over the course of the entire study. Study primary outcomes such as births in the last 30 days by site (Table 1) and total deliveries by site were also included. Lastly, several plots were available to visualize these figures such as participant accrual and total deliveries (outcomes), overall and by site, and the accrual prediction plot (Figure 2). Accrual reports were generated using R (version 3.6.2; The R Foundation) and included a description of utilized packages. Review of this weekly report became a primary feature of discussion among the team. It kept the team focused and often pointed to early opportunities to adjust recruitment and follow-up tactics to stay on track.

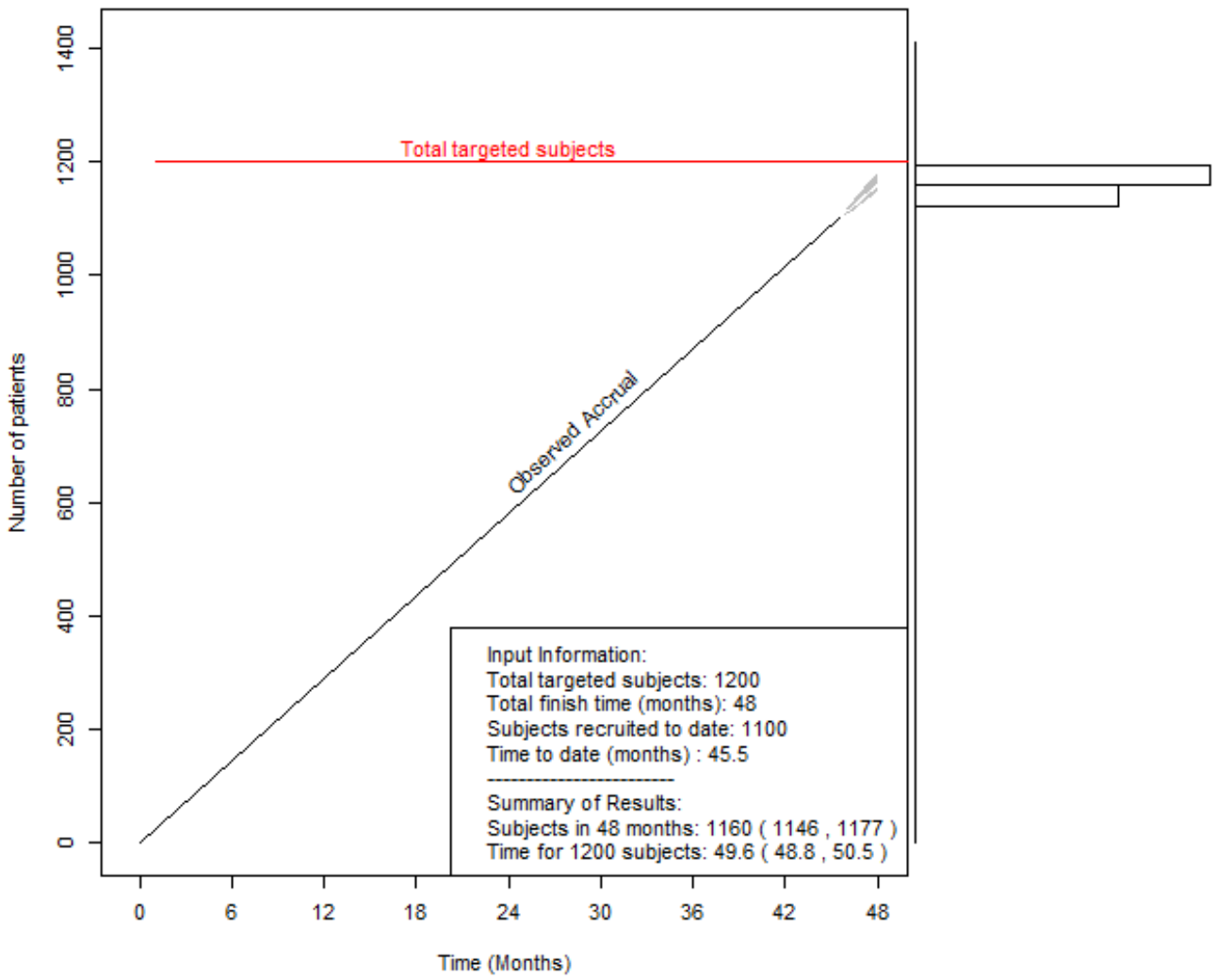
**Table 1.** Accrual by site in the current month, the past 30 days, and overall.

| Site | Participants, n |
|---|---|
| **Current month** | |
| Ohio State University | 5 |
| University of Kansas Medical Center | 4 |
| **Last 30 days** | |
| Ohio State University | 7 |
| University of Kansas Medical Center | 10 |
| **Total accrual** | |
| Ohio State University | 359 |
| University of Kansas Medical Center | 489 |
| University of Cincinnati | 252 |



XSL•FO
RenderX

**Figure 2.** Accrual prediction plot; accrual prediction with visualization plots to demonstrate the predicted completion date with a 95% prediction interval and the posterior predictive distribution. Redline represents the study deadline. The black line represents the current accrual rate.




## Study Protocol Adherence Reports

An AEs report was generated twice weekly (and immediately for serious AEs) for principal investigators to determine attribution and relatedness. Prompt review of AEs by the principal investigator kept the trial in regulatory compliance and allowed the investigators to quickly see any safety or data entry issues. A principal investigator sign off report that was generated twice monthly included participants who had completed the study and allowed their records to be reviewed for accuracy in a timely manner.

A resupply request report was generated for study coordinators and pharmacists. Resupply request reports detailed upcoming medication refills for study participants (Figure 3). For each study participant, her last medication refill date was recorded along with a date 2 weeks prior to when the current prescription was expected to run out. This allowed time for the prescription to be filled and mailed to the participant and prevented delays in treatment. Study coordinators at all sites benefited from reviewing this report and to anticipate which participants required telephonic or in-person follow-up to ensure refills arrived in good condition.

The delivery watch list report was generated weekly for study teams at each site and made available to all members of the study team. The report included the estimated delivery date for participants who were due to deliver in the next couple of months. The report kept teams aware of participants to watch for in the delivery service so that the collection of necessary delivery blood samples could be ensured.



XSL•FO
RenderX

**Figure 3.** Automated weekly pharmacy refill report. DHA: docosahexaenoic acid

07:00 Thursday, January 18, 2018

ADORE TRIAL

***Refill Request***

| Study ID | Treatment | Language | Status | EDD | Empty Date | Refill Num | Labelled |
|---|---|---|---|---|---|---|---|
| | Solid Blue | English | Active/On Treatment | 02/17/2018 | 01/27/2018 | 4 | No |
| | Black Stripe | English | Active/On Treatment | 02/18/2018 | 01/27/2018 | 3 | Yes |
| | Black Stripe | Spanish | Active/On Treatment | 02/18/2018 | 01/27/2018 | 3 | Yes |
| | Solid Blue | English | Active/On Treatment | 03/03/2018 | 01/27/2018 | 3 | Yes |
| | Solid Blue | Spanish | Active/On Treatment | 02/26/2018 | 01/27/2018 | 3 | Yes |
| | Black Stripe | English | Active/On Treatment | 03/30/2018 | 01/27/2018 | 2 | No |
| | Solid Blue | English | Active/On Treatment | 04/21/2018 | 01/27/2018 | 2 | No |
| | Black Stripe | Spanish | Active/On Treatment | 04/29/2018 | 01/23/2018 | 1 | No |
| | Black Stripe | English | Active/On Treatment | 06/15/2018 | 01/25/2018 | 1 | No |
| | Black Stripe | English | Active/On Treatment | 06/14/2018 | 01/26/2018 | 1 | No |
| | Black Stripe | English | Enrolled | 05/20/2018 | 01/26/2018 | 1 | No |
| | Black Stripe | English | Active/On Treatment | 05/27/2018 | 01/26/2018 | 1 | No |
| | Black Stripe | English | Active/On Treatment | 02/18/2018 | 01/27/2018 | 4 | No |
| | Black Stripe | English | Active/On Treatment | 01/25/2018 | 01/27/2018 | 4 | No |
| | Black Stripe | English | Active/On Treatment | 02/08/2018 | 01/27/2018 | 3 | Yes |
| | Solid Blue | English | Active/On Treatment | 03/22/2018 | 01/27/2018 | 2 | No |
| | Black Stripe | English | Active/On Treatment | 05/17/2018 | 01/26/2018 | 1 | No |
| | Black Stripe | English | Active/On Treatment | 04/30/2018 | 01/27/2018 | 1 | No |
| | Black Stripe | English | Active/On Treatment | 05/05/2018 | 01/30/2018 | 1 | No |
| | Solid Blue | English | Active/On Treatment | 05/25/2018 | 01/31/2018 | 1 | No |
| | Black Stripe | English | Active/On Treatment | 05/26/2018 | 02/01/2018 | 1 | No |
| | Black Stripe | English | Active/On Treatment | 02/18/2018 | 01/27/2018 | 3 | Yes |
| | Black Stripe | English | Active/On Treatment | 03/28/2018 | 01/27/2018 | 2 | No |

***Labelled column No means don't send 200 mg DHA; Labelled column Yes means send 200 mg DHA***

## Data Validation Reports

Data query reports were generated weekly for each site and identified missing or invalid data on the eCRF so they could be corrected by study teams in the CRIS. As mentioned previously, electronic data capture produces fewer errors and allows missing or invalid data to be quickly and automatically identified [14]. The Data Query reports used in this study leveraged the accessibility of electronically captured data to automate data review processes for significant time-saving during data harmonization.

## Interim Analysis

Eleven interim analyses were conducted during the ADORE trial during which treatment randomization was adjusted in accordance with the Bayesian Adaptive Design. An average of 3-4 days were spent on each interim analysis. Of these days,



XSL•FO
RenderX

two of them were working days during which the analyst locked the data, reviewed the data, and updated the randomization table.

### Measures of Benefit

Two primary measures of benefit for the ADORE trial were captured: protocol deviation and loss to follow-up. Protocol deviation is a common occurrence in clinical trials, with an average reporting rate of 0.08 protocol deviations per participant [17]. For 1100 participants, there were 2 protocol deviations in the ADORE trial, constituting a rate of 0.0018 protocol deviations per participant. Of the 1100 participants in the ADORE trial, 68 were lost to follow-up. This constituted 6% of the study participants and was lower than the predicted loss of 15%.

The effect of these systems on study management was to allow a seamless process of data entry, feedback, and communication between study sites. The process of data capture, validation, flagging for inconsistencies or errors, and data correction occurred automatically within a single system. The statistical analysis team at the central site could track each of these processes as they occurred and communicate with study teams completely within CRIS. This eliminated the need to interact with multiple systems for these processes and allowed the CRIS to be used as an encompassing clinical trial data tool.

## Discussion

### Principal Findings

Centralized data capture can be more cost effective, accurate, and efficient than decentralized systems when implemented effectively [18,19]. However, the variability in how centralized data capture is performed presents an operational hurdle for effective implementation. If mismanaged, centralized data capture can exacerbate the complexity that arises from differences between study sites [20]. CTMSs can make centralized data capture more feasible through the availability of key features for multisite management. Unfortunately, many of these systems are either quite expensive or lack the functionality to manage large trials effectively [21].

The primary aim of developing the KUMC CRIS was to design an adaptable CTMS that streamlined the centralized data management process through automated monitoring, verification, integration, and reporting. The ADORE trial presented a unique opportunity through which to test the benefits and limitations of this system for future use. Throughout the ADORE trial, the CRIS was able to ensure the quality of the study data through frequent data quality checks, regularly generated operation reports with little oversight, and ensuring familiarity with CRIS was consistent across systems. One crucial aspect of CRIS implementation in this trial was ensuring that a member from each site travelled to KUMC for training specifically in using this system. This ensured that at least one member of each study team had direct contact with the central data management team, and the CRIS before their site's study team was trained.

Other web-based CTMSs have been designed for multisite trials, and these systems served as a useful reference to build from when deciding what functionality should be included in the CRIS [22,23]. Durkalski et al [22] detailed a clinical trial management system used in a multisite trial that included functions such as centralized participant enrollment and randomization, real-time reporting of CRF completion rates, and real-time data validation upon data entry. The CRIS system was able to integrate these features as well as accrual tracking and prediction, patient assignment and randomization using the Bayesian Adaptive Design, and centralized training tracking for study staff. One of the chief benefits of sharing the design and structure of these novel systems is that it allows other research teams to begin with a reference for essential functionality and build on that design even further. Development of the CRIS through utilization in other clinical trials will allow us to discover what additional functionality is needed. Further, this study could help other research centers adapt and improve our CRIS design. Thus, large research centers that frequently run multisite trials can help each other become more efficient.

It is important to emphasize that the tools developed for the CRIS were designed to facilitate strong teamwork and communication between teams at all sites. With limited resources at each site, a centralized data management design can allow multisite trials to allocate more resources toward effective trial execution. We would recommend the implementation of similar tools for any research center that frequently acts as a primary site in multisite trials.

Future developments of the CRIS will occur through implementation in subsequent KUMC multisite trials. Aspects of the CRIS such as operational reports and trial accrual reports will be configured for each of these trials in accordance with their unique circumstances. These configured designs can then be used as templates for future trials with similar features, eventually resulting in a library of potential reports that can easily be tailored to each new trial.

As the costs of clinical trials continue to rise, new techniques will be required to manage resources effectively [24]. Multisite trials utilizing centralized data capture are cost-effective because they remove the requirement of data management teams at each site [25]. These designs can, however, present significant challenges to data management and communication between study teams. Several tools can be used to address these challenges, such as automated operational reports, accrual reporting, consistent training, and centralized data validation. Thus, the development of a CTMS specifically for this purpose can provide significant cost savings and efficient trial execution for large research centers. Final data analysis and unmasking randomization took couple of days instead of weeks or months, which is a testament of all the hard work that has been vested by the team from day 1.

### Limitations

While the CRIS system was able to be effectively utilized in the ADORE trial, there was a significant investment required for both development and onboarding.

First, the time required to build the database within CRIS with minimal customization is 3-4 weeks for phase 1, 2, or 3 studies. The database design and data validation systems could also require more time for development if the study protocol is not



XSL•FO
RenderX

finalized. This development time may be alleviated as the system is used for more studies and we are able to develop templates for common study design features. At the present, however, this represents a significant time investment. Furthermore, the customization of eCRFs was limited only to the central data management team. The second limitation of this system was the process of standardized training and onboarding. This required a time and labor investment on the part of study teams, and a willingness to learn how to use the CRIS systems. As CRIS is updated with subsequent trials, the training material will also need to be updated. Because study personnel at each site need to be trained on the same material for standardization purposes, this could require additional time and resources.

## Conclusions

This study shows that multicenter trial success is dependent on prompt orchestration. Utilizing a platform to ease the execution steps is crucial, which the team at KUMC has demonstrated through the use of the CRIS.

## Acknowledgments

We are grateful for the support of numerous study personnel who were responsible for recruiting participants at all 3 sites, communicating with them on a monthly basis, and collecting the data critical for the study. They did so with diligence and care, and this study would not have been possible without them or the nurses at each medical center who took on the additional responsibilities of obtaining blood samples and informing staff of their availability. We are very grateful as well to our Data Safety and Monitoring Board (DSMB), excellently chaired by Daniel Robinson, MD, of Lurie Children's Hospital in Chicago, Arthur Evans, MD, of the University of Cincinnati, Ian Griffin, MD, of the Biomedical Research Institute of New Jersey, and Ardythe Morrow, PhD, of the University of Cincinnati. Alexander Alsup Graduate Student in Department of Biostatistics & Data Science. As a group, they were all engaged and invaluable in helping us refine our system for reporting adverse events. We are also grateful to the 1100 women who enrolled in the study.

## Authors' Contributions

DPM oversaw all aspects of drafting, revision, and final approval of the manuscript. SEC, BJG, and CJV were principal investigators and designed the study with input from all other authors. EHK developed the manual of operating procedures, oversaw conduct of study operation at all sites. ARB and DPM were responsible for verifying the data within Velos. DPM developed the electronic records for data entry in conjunction with EHK and ARB. DPM, EHK, and ARB wrote the manuscript, but all authors contributed their insights to the final version.

## Conflicts of Interest

SEC has received honorariums for presentations about DHA in infancy and pregnancy. CJV is an employee of RB Nutrition, which produces infant formulas and supplements; however, RB had no involvement in the study execution or analysis. She conducted this study through her role as an adjunct professor at the University of Cincinnati. The other authors have no competing interests.

Multimedia Appendix 1
List of Case Report Forms used for ADORE study.
[DOCX File , 51 KB - formative_v5i12e30368_app1.docx ]

## References

1. Logviss K, Krievins D, Purvina S. Characteristics of clinical trials in rare vs. common diseases: A register-based Latvian study. PLoS One 2018;13(4):e0194494 [FREE Full text] [doi: 10.1371/journal.pone.0194494] [Medline: 29614103]
2. Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, et al. Conducting multicenter research in healthcare simulation: Lessons learned from the INSPIRE network. Adv Simul (Lond) 2017;2:6 [FREE Full text] [doi: 10.1186/s41077-017-0039-0] [Medline: 29450007]
3. Youssef NCF, Reinhart K, Sakr Y. The pros and cons of multicentre studies. Neth J Crit Care 2008;12(3):120.
4. Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2). International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). 2016 Nov 09. URL: https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf [accessed 2021-12-09]
5. Forjuoh SN, Helduser JW, Bolin JN, Ory MG. Challenges Associated with Multi-institutional Multi-site Clinical Trial Collaborations: Lessons from a Diabetes Self-Management Interventions Study in Primary Care. J Clin Trials 2015;05(03):219. [doi: 10.4172/2167-0870.1000219]
6. Spirito A, Abebe KZ, Iyengar S, Brent D, Vitiello B, Clarke G, et al. Sources of site differences in the efficacy of a multisite clinical trial: the Treatment of SSRI-Resistant Depression in Adolescents. J Consult Clin Psychol 2009 Jun;77(3):439-450 [FREE Full text] [doi: 10.1037/a0014834] [Medline: 19485586]



XSL•FO
RenderX

7. Rahbar MH, Lee M, Hessabi M, Tahanan A, Brown MA, Learch TJ, et al. Harmonization, data management, and statistical issues related to prospective multicenter studies in Ankylosing spondylitis (AS): Experience from the Prospective Study Of Ankylosing Spondylitis (PSOAS) cohort. Contemp Clin Trials Commun 2018 Sep;11:127-135 [FREE Full text] [doi: 10.1016/j.conctc.2018.07.004] [Medline: 30094388]
8. Messerer D, Porzsolt F, Hasford J, Neiss A. [Advantages and problems of multicenter therapy studies exemplified by a study of the treatment of metastasizing renal cell carcinoma with recombinant interferon-alpha-2c]. Onkologie 1987 Feb;10(1):43-49. [doi: 10.1159/000216367] [Medline: 2954005]
9. Kahan BC, Harhay MO. Many multicenter trials had few events per center, requiring analysis via random-effects models or GEEs. J Clin Epidemiol 2015 Dec;68(12):1504-1511 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.03.016] [Medline: 25985892]
10. Carlson SE, Gajewski BJ, Valentine CJ, Rogers LK, Weiner CP, DeFranco EA, et al. Assessment of DHA on reducing early preterm birth: the ADORE randomized controlled trial protocol. BMC Pregnancy Childbirth 2017 Feb 13;17(1):62 [FREE Full text] [doi: 10.1186/s12884-017-1244-5] [Medline: 28193189]
11. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
12. Code of Federal Regulations: A point in time eCFR system. National Archives and Records Administration. URL: https://www.ecfr.gov/ [accessed 2021-12-09]
13. Fleischmann R, Decker A, Kraft A, Mai K, Schmidt S. Mobile electronic versus paper case report forms in clinical trials: a randomized controlled trial. BMC Med Res Methodol 2017 Dec 01;17(1):153 [FREE Full text] [doi: 10.1186/s12874-017-0429-y] [Medline: 29191176]
14. Walther B, Hossin S, Townend J, Abernethy N, Parker D, Jeffries D. Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. PLoS One 2011;6(9):e25348. [doi: 10.1371/journal.pone.0025348] [Medline: 21966505]
15. Brown AR, Gajewski BJ, Aaronson LS, Mudaranthakam DP, Hunt SL, Berry SM, et al. A Bayesian comparative effectiveness trial in action: developing a platform for multisite study adaptive randomization. Trials 2016 Aug 31;17(1):428 [FREE Full text] [doi: 10.1186/s13063-016-1544-5] [Medline: 27577191]
16. Haidich A, Ioannidis JP. The Gini coefficient as a measure for understanding accrual inequalities in multicenter clinical studies. J Clin Epidemiol 2004 Apr;57(4):341-348. [doi: 10.1016/j.jclinepi.2003.09.011] [Medline: 15135834]
17. Jalgaonkar SV, Bhide SS, Tripathi RK, Shetty YC, Marathe PA, Katkar J, et al. An Audit of Protocol Deviations Submitted to an Institutional Ethics Committee of a Tertiary Care Hospital. PLoS One 2016;11(1):e0146334 [FREE Full text] [doi: 10.1371/journal.pone.0146334] [Medline: 26735850]
18. Welker JA. Implementation of electronic data capture systems: barriers and solutions. Contemp Clin Trials 2007 May;28(3):329-336. [doi: 10.1016/j.cct.2007.01.001] [Medline: 17287151]
19. Walden A, Nahm M, Barnett ME, Conde JG, Dent A, Fadiel A, et al. Economic analysis of centralized vs. decentralized electronic data capture in multi-center clinical studies. Stud Health Technol Inform 2011;164:82-88 [FREE Full text] [Medline: 21335692]
20. Park YR, Yoon YJ, Koo H, Yoo S, Choi C, Beck S, et al. Utilization of a Clinical Trial Management System for the Whole Clinical Trial Process as an Integrated Database: System Development. J Med Internet Res 2018 Apr 24;20(4):e103 [FREE Full text] [doi: 10.2196/jmir.9312] [Medline: 29691212]
21. Leroux H, McBride S, Gibson S. On selecting a clinical trial management system for large scale, multi-centre, multi-modal clinical research study. Stud Health Technol Inform 2011;168:89-95. [Medline: 21893916]
22. Durkalski V, Wenle Z, Dillon C, Kim J. A web-based clinical trial management system for a sham-controlled multicenter clinical trial in depression. Clin Trials 2010 Apr;7(2):174-182. [doi: 10.1177/1740774509358748] [Medline: 20083496]
23. Unützer J, Choi Y, Cook IA, Oishi S. A web-based data management system to improve care for depression in a multicenter clinical trial. Psychiatr Serv 2002 Jun;53(6):671-3, 678. [doi: 10.1176/ps.53.6.671] [Medline: 12045303]
24. Park J, Moon SJ, Lee J, Jeon J, Jang K, Kim M. The first step to the powers for clinical trials: a survey on the current and future Clinical Trial Management System. Transl Clin Pharmacol 2018 Jun;26(2):86-92 [FREE Full text] [doi: 10.12793/tcp.2018.26.2.86] [Medline: 32055554]
25. Collier R. Rapidly rising clinical trial costs worry researchers. CMAJ 2009 Feb 03;180(3):277-278 [FREE Full text] [doi: 10.1503/cmaj.082041] [Medline: 19188619]

## Abbreviations

**AE:** adverse event
**CRIS:** Comprehensive Research Information System
**CTMS:** clinical trial management system
**DHA:** docosahexaenoic acid
**DSMB:** Data Safety and Monitoring Board



XSL•FO
RenderX

**eCRF:** electronic case report form
**ICH GCP:** International Council for Harmonization Guideline for Good Clinical Practice
**KUMC:** University of Kansas Medical Center
**RCT:** randomized controlled trial
**WCG:** Western Institutional Review Board–Copernicus Group, Inc

*Edited by G Eysenbach; submitted 11.05.21; peer-reviewed by K Zhai, S Yeatts; comments to author 06.08.21; revised version received 30.09.21; accepted 21.11.21; published 22.12.21.*

*Please cite as:*
*Mudaranthakam DP, Brown A, Kerling E, Carlson SE, Valentine CJ, Gajewski B*
*The Successful Synchronized Orchestration of an Investigator-Initiated Multicenter Trial Using a Clinical Trial Management System and Team Approach: Design and Utility Study*
*JMIR Form Res 2021;5(12):e30368*
*URL: https://formative.jmir.org/2021/12/e30368*
*doi:10.2196/30368*
*PMID:34941552*

©Dinesh Pal Mudaranthakam, Alexandra Brown, Elizabeth Kerling, Susan E Carlson, Christina J Valentine, Byron Gajewski. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Understanding Adolescents' Perceptions and Aspirations Towards Their Relationship With Personal Technology: Survey Study

Muhammad Jee[1*]; Alsa Khan[1*]; Nazneen Nazneen[2], PhD

[1]Twinbits, Flower Mound, TX, United States

[2]Userwise Consulting, Mountain View, CA, United States

[*]these authors contributed equally

**Corresponding Author:**
Muhammad Jee
Twinbits
1201 Big Falls Dr
Flower Mound, TX, 75028
United States
Phone: 1 214 972 8988
Email: muhammadjee1@gmail.com

## Abstract

**Background:** Understanding adolescents' relationship with technology is a pressing topic in this digital era. There seem to be both beneficial and detrimental implications that originate from use of technology by adolescents. Approximately 95% of adolescents have access to a smartphone, and several studies show a positive correlation between screen addiction and trends of anxiety and depression. At the same time, research shows that two-thirds of adolescents believe that technology is a necessity for connecting and making new friends.

**Objective:** The aim of this formative study was to understand adolescents' perception of their own and others' relationship with personal technology.

**Methods:** A survey was conducted with 619 adolescents ranging in age from 13 to 19 years. Adolescents were asked how they perceived the relationship with their personal technology, how they perceived others' (parents, siblings, or friends) relationship with personal technology, and how they wish to relate to their personal technology in the future.

**Results:** "Essential," "Distractive," and "Addictive" were the most commonly selected descriptors to describe both adolescents' own relationship with technology (essential: 106/619, 17.1%; distractive: 105/619, 17%; addictive: 88/619, 14.2% ) and others' relationship as well (essential: 96/619, 15.6%; distractive: 88/619, 14.3%; addictive: 90/619, 14.5%). Adolescents selected "Provides an escape" more to describe their own relationship with technology. Whereas, they selected "It's just a tool" and "Creates Barrier" more to describe others' relationship with technology. These trends are consistent across ages and genders. In addition, adolescents' aspirations for their relationship with their personal technology varied across ages: 13 to 15-year olds' top choice was "best friend", 16 to 17-year olds' top choice was "I don't believe in personal connection with mobile technology," and 18 to 19-year olds' top choice was "My personal assistant."

**Conclusions:** Our 3-lens method allows us to examine how adolescents perceive their relationship with personal technology in comparison to others, as well as their future technological aspirations. Our findings suggest that adolescents see both communalities as well as differences in their own and others' relationships with technology. Their future aspirations for personal technology vary across age and gender. These preliminary findings will be examined further in our follow-up research.

***(JMIR Form Res 2021;5(12):e27852)*** doi:10.2196/27852

**KEYWORDS**

adolescents' perceptions; personal technology; technology relationship; adolescents as technology users; adolescents as technology bystanders



XSL•FO
RenderX

## Introduction

### Background

As personal technology is advancing rapidly, adolescents' technology adoption and usage are continuously on the rise. Smartphone use in the adolescent population is nearly universal. Approximately 95% of adolescents in the USA say they have a smartphone or have access to one [1].

Smartphones provide a number of benefits, including increased productivity, efficient information seeking, and enhanced access to meaningful support networks during times of distress and ongoing illnesses [2-4]. Nearly two-thirds of the US adolescent population report that personal technology and social media enable them to make new friends and allow them to bond with people [5]. Youth social media use is also related to an increase in empathy; both in their ability to understand (cognitive empathy) and share the feelings of peers (affective empathy) [6].

Despite these benefits, an increasing amount of studies have revealed the adverse effects associated with adolescents' hyperconnected relationship with personal technology. Consequences of excessive technology dependency include lower self-motivation, decreased social skills, screen addiction, and mental health problems such as anxiety and depression [1,7-9].

With technology having both positive and negative implications on adolescents, an in-depth understanding of adolescents' relationship with technology is necessary.

Research shows that due to unique social, cultural, and developmental factors, adolescents' technology use behaviors differ from those of adults [10-13]. Thus, adolescents' perceptions and choices about technology may not be identical to adults' and are worth investigating.

The goal of our study is to explore how adolescents perceive their own and others' relationship with technology. Understanding and acknowledging independent perceptions about personal technology is critical for informing technology use management, policies, design, health interventions, teaching, and parenting that resonate with adolescents' attitudes.

The key contribution of this paper is our holistic approach that interprets adolescents' relationship with personal technology via 3 lenses. Lens 1 is adolescents' perception of their current relationship with technology as users. Lens 2 is adolescents' perception of others' (parents, siblings, or friends) relationship with technology as bystanders. Lens 3 is adolescents' aspiration of what relationship, if any, they wish to have with their personal technology in the future. Unlike past research, our approach not only characterizes adolescents' perceptions of their own relationship with technology, but it also frames how adolescents perceive others' relationships with technology. Even though a few studies have also considered adolescents' perception of their parents' or peers' use of technology [14,15], our study examines a broader trusted social circle of parents, siblings, or friends. This 3-lens approach is a holistic method as it not only reflects how adolescents relate to technology presently but also covers future aspirations of how adolescents wish to relate to technology. Our 3-lens approach allows for an overarching comparison of adolescents' perception about relationship with technology as users versus as bystanders, as well as their current versus aspirational perspectives. Hence, our research identifies the similarities and differences in how adolescents perceive their own and others' relationships with technology.

In sum, this paper explores these open questions: How do adolescents perceive others' (parents, siblings, or friends) relationship with personal technology and how does it compare from theirs? What relationship, if any, do adolescents wish to have with their personal technology in the future and how does it relate to their perception of their present relationship with technology?

### Prior Work

#### *Psychological and Behavioral Implications of Technology*

The ubiquity of the smartphone has led researchers to devote much attention to the psychological, behavioral, and social implications of personal technologies. There is vast research into the benefits of social media: sustaining close friendships, building new connections with individuals from diverse backgrounds, demonstrating support for meaningful causes, and becoming civic minded [16,17]. Moreover, smartphone-based interventions enhance the effects of policies on a range of outcomes, including the adoption of positive healthy habits and educational activities [18,19]. Smartphone use seems to have significant impacts on improving student performance, teaching, and learning experiences and is regarded as a key component in the development of social environment [20]. Students benefit from the incorporation of smartphone use in educational activities by efficiently accessing vast course content, participating in debate sessions with professors, and retrieving information regarding student performance [21,22].

Young adults' excessive, addictive, and problematic technology use has been continuously reported alongside the rise in technology adoption. An abundance of research has correlated smartphone usage with negative psychological and behavioral implications, including anxiety, insomnia, and depression [23,24]. Constantly checking communication updates, feeling restless without close proximity to a phone, and suffering delays in professional performance due to prolonged phone activities are indicators of smartphone addiction [25,26]. Excessive use of smartphones can also result in various impacts on physical health, such as fatigue, indigestion, sleep issues, and eyesight problems [27,28]. Parasuraman et al [29] investigated the impact on daily life of smartphones on 55-year-old and 18-year-old age groups and learned that a significant portion of the participants had an addiction to smartphone usage but were unaware of it due to smartphones having become an integral part of their lives [29].

Adolescents who are addicted to personal technology face diminishing social skills and challenges in developing friendships in the real world [30]. Exploring the relationship between social internet use and loneliness, Nowland et al [31] concluded that loneliness can be reduced when the digital world



XSL•FO
RenderX

is used to maintain or forge more social connections. However, when people use the internet to avoid physical social activities and day-to-day problems, their loneliness increases. Consequently, they develop a preference to continue using the internet in a way that displaces time spent in offline social interactions. One study demonstrated that adolescents who were more addicted to their phones had an increased risk of feeling the 6 variants of social alienation: powerlessness, normlessness, meaninglessness, self-estrangement, cultural estrangement, and social isolation [32].

### *Correlation of Age and Gender With Technology Usage*

Many studies have also assessed the potential correlation of age, gender, and personality variables on smartphone addiction. Adolescents with somatization, poor self-control, interpersonal sensitivity, and hostility tend to be more likely to get addicted to smartphones [33-35]. Neuroticism, conscientiousness, and openness are traits that were found to be negatively correlated with smartphone addiction [36]. Furthermore, research shows that males and females use their phones for different reasons. Female smartphone use is more strongly related to sociability, interpersonal relationships, and the desire to maintain connections [7]. Comparatively, males use their smartphones more extensively for media sharing, video games, and online searches [36]. Many studies found smartphone addiction being more prevalent in younger adolescents [37,38]. A study involving 1529 students aged 11 to 18 years found that younger adolescents (11-14 years) had a higher prevalence of problematic smartphone usage than did older adolescents (15-18 years) [15].

### *Perceptions About Self and Others' Technology Usage*

When US adolescents were asked directly, 31% perceived the effects of social media as mostly positive, 45% believed the effects to be neither positive nor negative, and 24% stated the effects as mostly negative. Those who considered the effects of social media to be positive indicated that it helps in maintaining connectivity with family and friends, obtaining access to information, and meeting like-minded people. Those who considered the effects of social media to be negative explained that it increases the risks of addiction, hate speech, neglecting face-to-face contacts, and obtaining unrealistic views of others' lives [39]. Similarly, Ozkan and Solmaz [40] examined the correlation between mobile addiction and perception of one's own personal technology. Through a survey of 18 to 23-year-old university students, they concluded that addiction to technology is related to self-perception technology use. There was a statistically significant correlation between the perception that smartphone apps are useful tools for communicating with people and time being spent on the phone.

Davis and Dinhopl [14] studied the similarities and differences in the way parents and adolescents described their own and each other's phone use in the context of family life. Both expressed a lack of agency in their own and each other's smartphone use, feeling displaced by the other's smartphone and highly reliant on their own smartphone. In addition, parents felt guilty about the impact of their phone overuse on their children, whereas adolescents' expression of guilt was based on what their parents and society thought of their phone use. Lopez-Fernandez et al [15] conducted a study in which adolescents were asked to indicate whether they felt that any of their peers used their smartphones excessively. Results showed that adolescents with problematic smartphone usage were more likely to consider their peers' smartphone usage to be problematic.

Previous research has investigated adolescents' relationship with technology through their use of technology, their own perceptions, or their parents' perspectives. Whereas, our research used a multi-lens approach to explore alignments and mismatches in adolescent's perceptions of their own versus others' relationship with technology as well as their current versus aspirational relationship.

## Methods

### Participants

A total of 619 adolescents participated in the online survey. Among the participants, 58.8% (364/619) were females, while 39.9% (247/619) were males. Table 1 shows participants' demographics.

From March 2020 to April 2020, this online survey was deployed through different social media platforms to a diverse range of adolescents. Participation in the survey was voluntary, and no incentives were offered. In addition, no identifiable data were collected.

**Table 1.** Participants' demographics (N=619).

| Demographics | Segment size, n (%) |
|---|---|
| **Gender** | |
| Male | 247 (39.90%) |
| Female | 364 (58.80%) |
| Prefer not to specify | 8 (1.29%) |
| **Age** | |
| 13-15 years | 76 (12.28%) |
| 16-17 years | 410 (66.24%) |
| 18-19 years | 133 (21.49%) |



XSL•FO
RenderX

## Three-Lens Approach

The survey included 3 main questions representing 3 lenses to understand how adolescents relate to technology. These questions are designed to analyze adolescents' perceptions of their own personal technology (lens 1: adolescents as users) versus the opinion of adolescents of their parents, siblings, and friends' use (lens 2: adolescents as bystanders) and to identify the type of relationship they desire to have with their technology (lens 3: aspirational). Figure 1 illustrates our novel 3-lens approach.

By bringing all 3 lenses together, we captured a more holistic understanding of adolescents' perception of their own relationship with technology in comparison with both adolescents' perceptions of others' relationship with technology as well as adolescents' aspirations of their technology relationship. Table 2 lists all 3 questions along with options presented in the survey.

**Figure 1.** A model illustrating our 3-lens approach.

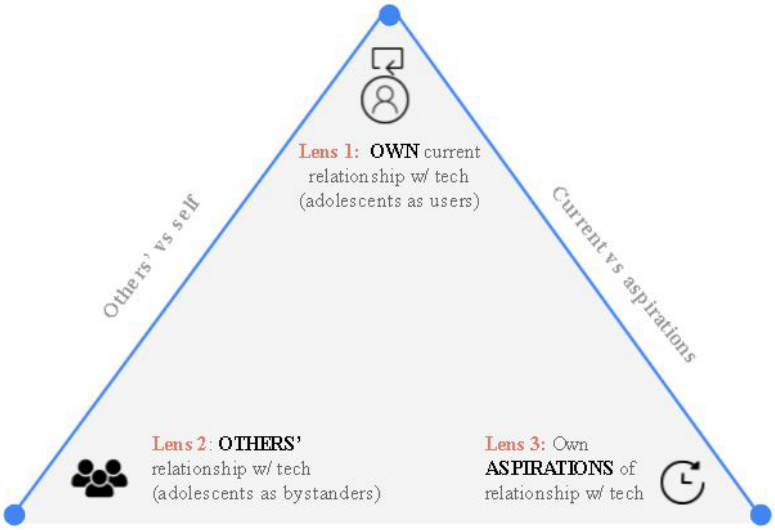



**Table 2.** List of all 3 questions along with options presented in the survey.

| Question | Options |
| --- | --- |
| Q1: Choose 3 options that best describe the relationship between you and your personal technology. | Options: Creates barrier, Enabler, Overwhelming, It's just a tool, Empowering, Private, Addictive, Essential, Distractive, Hurtful, Provides an escape, Emotional, Other |
| Q2: Choose 3 options that you think best describe the relationship between your parents, siblings, or friends and their personal technology. | Options: Creates barrier, Enabler, Overwhelming, It's just a tool, Empowering, Private, Addictive, Essential, Distractive, Hurtful, Provides an escape, Emotional, Other |
| Q3: What persona would you wish your personal technology to match the most with? | Options: I don't believe in personal connection with mobile technology, My best friend, My twin sibling, My personal assistant, My coach/mentor, Other |

## Descriptor Selection Process

In question 1, out of 12 descriptors, respondents were asked to choose 3 descriptors that best describes their relationship with their personal technology. Similarly, in question 2, from the same 12 descriptors, adolescents were asked to choose 3 that best described their perception of the relationship of their parents, siblings, or friends with their personal mobile technology.

For context, 2 of the 3 authors (AK and MH), who are high school students and adolescents, had firsthand experience learning from their peers and immersing in the high school environment and other adolescents' online social networks. The level of communication, comfort, and openness at the peer level, provided a unique platform to this research.

In round 1, an initial list of descriptors for questions 1 and 2 was compiled based on learning from unstructured observations and semistructured interviews of 5 high school peers. In the interviews, the adolescents were asked an open-ended question to identify descriptors they would use to describe relationships with personal technology. The descriptors that were mentioned were "Empowering," "Provides an escape," "It's just a tool," and "Enabler." These 4 words were used as a general guideline to develop a larger list of descriptors.

In round 2, we added 8 additional descriptors to the initial list by further learning from literature reviews related to adolescents' usage of technology, and personality and profile assessments like the Activity Vector Analysis and Adjective Check List [3-5,7-9,23,24,29,32,41-43]. Out of a list of positive and negative adjectives, these profile assessments allow individuals to choose adjectives that best describe themselves and others' perception of them.

A list of 12 descriptors was finalized by ensuring that descriptors fall within a spectrum of positive and negative for a diverse understanding of adolescents and technology relationships. After the list was completed, we conducted another round of



XSL•FO
RenderX

semistructured interviews with a new sample of 4 high school adolescents to judge each descriptor as a "positive," "neutral," and "negative" relationship with personal technology. The descriptors provided in the survey as response to questions 1 and 2 were categorized by words ranging from those that describe the strongest positive relationships (Essential, Empowering, Enabler, Private, Emotional), the neutral relationships (It's just a tool), and the strongest negative relationships (Provides an escape, Distractive, Creates Barrier, Overwhelming, Hurtful).

Options for question 3, in which adolescents were asked which persona they would wish for in their personal technology, were based on trusted relationships (eg, best friends, siblings), exciting technology offerings (eg, personal assistant Siri, Alexa, Google Assistant), and coach or mentors (eg, learning and fitness apps). The option "I don't believe in personal connection with mobile technology" was included for completeness. Selections for all 3 questions were randomized to eliminate any order effect.

# Results

## Analysis

This section summarizes analysis of key responses within individual lenses.

### *Adolescents' Perception of Their Own Relationship With Their Personal Technology as Users*

The top 3 selected descriptors that adolescents chose to describe their own relationship with technology were "Distractive," "Essential," and "Provides an escape." This trend was consistent across genders and age groups (13 to 15-, 16 to 17-, and 18 to 19-year olds).

However, "Distractive" was chosen more often by females. In addition, 13 to 15-year olds selected "Provides an escape" the most. Among all descriptors, "Hurtful" was the least selected across all ages and genders. Table 3 summarizes the percentage of times each descriptor was selected overall (by everyone) and within each segment (male, female, 13 to 15-year old, 16 to 17-year old, 18 to 19-year old).

**Table 3.** Table of the percentage of times each descriptor was selected for describing adolescents' own relationship with technology. Sorted in descending order of overall percentages. Absolute values are not provided as percentages have been adjusted by weighting. Participants were asked to choose 3 choices to describe their own relationship with technology; the values are normalized to sum up to 100%.

| Descriptors in Q1 | Overall | Male | Female | Age 13-15 | Age 16-17 | Age 18-19 |
|---|---|---|---|---|---|---|
| Distractive | 17% | 14% | 20% | 14% | 18% | 17% |
| Essential | 17% | 17% | 17% | 14% | 18% | 18% |
| Provides an escape | 16% | 15% | 17% | 19% | 16% | 15% |
| Addictive | 14% | 13% | 15% | 14% | 14% | 14% |
| Private | 8% | 8% | 8% | 10% | 8% | 7% |
| It's just a tool | 8% | 11% | 5% | 8% | 7% | 9% |
| Empowering | 7% | 8% | 6% | 6% | 6% | 7% |
| Enabler | 5% | 7% | 3% | 2% | 5% | 5% |
| Other | 4% | 5% | 2% | 11% | 2% | 4% |
| Overwhelming | 3% | 2% | 4% | 4% | 3% | 4% |
| Emotional | 3% | 3% | 3% | 4% | 3% | 2% |
| Creates barrier | 2% | 2% | 2% | 2% | 2% | 2% |
| Hurtful | 1% | 0% | 1% | 1% | 1% | 1% |

### *Adolescents' Perception of Others' Relationship With Personal Technology as Bystanders*

To describe others' (parents, siblings, friends) relationship with technology, the top 3 descriptors selected by adolescents were "Essential," "Distractive," and "Addictive." Among all descriptors, "Hurtful" was least selected. These trends were consistent across genders and age groups. Table 4 summarizes the percentage of times each descriptor was selected.



XSL•FO
RenderX

**Table 4.** Table of the percentage of times each descriptor was selected for describing others' relationship with technology. Sorted in descending order of overall percentages. Absolute values are not provided as percentages have been adjusted by weighting. Participants were asked to choose 3 choices to describe their own relationship with technology; the values are normalized to sum up to 100%.

| Descriptors in Q2 | Overall | Male | Female | Age: 13-15 | Age: 16-17 | Age: 18-19 |
|---|---|---|---|---|---|---|
| Essential | 16% | 15% | 16% | 14% | 16% | 15% |
| Distractive | 14% | 14% | 15% | 13% | 14% | 15% |
| Addictive | 14% | 13% | 15% | 13% | 15% | 15% |
| It's just a tool | 13% | 13% | 13% | 16% | 13% | 13% |
| Private | 9% | 9% | 8% | 8% | 9% | 8% |
| Provides an escape | 8% | 9% | 7% | 7% | 7% | 9% |
| Creates Barrier | 7% | 6% | 7% | 5% | 8% | 5% |
| Empowering | 5% | 6% | 4% | 7% | 4% | 5% |
| Enabler | 5% | 5% | 5% | 6% | 6% | 4% |
| Overwhelming | 4% | 4% | 4% | 4% | 4% | 4% |
| Emotional | 3% | 4% | 3% | 4% | 3% | 4% |
| Hurtful | 2% | 2% | 2% | 3% | 2% | 2% |
| Other | 2% | 1% | 1% | 5% | 1% | 2% |

### *Adolescents' Aspirations of a Persona They Wish for Their Personal Technology*

Adolescents' aspirations of a persona for their personal technology varied across age ranges. The majority (39/74, 52%) of females in the 13 to 15-year old group selected "My best friend" the most. The 16 to 17-year-old group selected "I don't believe in personal connection with mobile technology" and "My personal assistant", whereas the 18 to 19-year-old group chose "My personal assistant" most frequently. Table 5 shows the percentage of times each persona was selected.

**Table 5.** Table of the percentage of times each persona was selected as adolescents' aspiration of their personal technology.

| Q3 | Overall, n/N (%) | Age 13-15, n/N (%) | Age 16-17, n/N (%) | Age 18-19, n/N (%) |
|---|---|---|---|---|
| My personal assistant | 208/619 (33.6) | 11/76 (14.5) | 136/410 (33.2) | 61/133 (45.9) |
| I don't believe in personal connection with mobile technology | 199/619 (32.1) | 24/76 (31.6) | 141/410 (34.4) | 34/133 (25.5) |
| My best friend | 148/619 (23.9) | 31/76 (40.8) | 100/410 (24.4) | 17/133 (12.8) |
| My coach/mentor | 34/619 (5.5) | 6/76 (7.9) | 16/410 (3.9) | 12/133 (9.0) |
| My twin sibling | 19/619 (3.1) | 2/76 (2.6) | 11/410 (2.7) | 6/133 (4.5) |
| Other | 11/619 (1.8) | 2/76 (2.6) | 6/410 (1.5) | 3/133 (2.2) |

## Discussion

### Cross-comparison and Key Synthesis

This section addresses the research questions stated in the introduction through the cross-comparison of lenses.

### *Understanding Adolescents' Perception of Own Versus Others' Relationship With Technology*

A side-by-side comparison of adolescents' own relationship (lens 1) versus others' relationship (lens 2) revealed some notable similarities in the way adolescents perceive their own versus others' relationship with personal technology. The Venn diagram in Figure 2 shows an overview of the alignment and mismatch between adolescents' perceptions of their own versus others' relationship with personal technology.

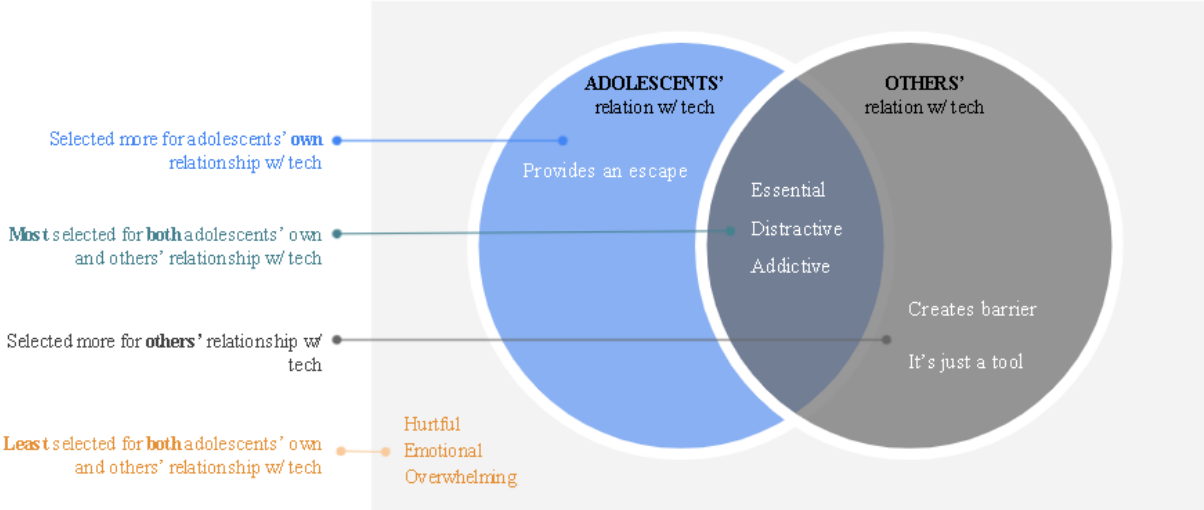
We found that adolescents perceive their own as well as others' relationship with personal technology as "Essential," Distractive," and "Addictive." However, there were some key differences in how adolescents perceive their own versus others' relationship with their personal technology. Adolescents associate "Provide an escape" more with their own relationship with technology. However, they attributed "It's just a tool" and "Creates Barrier" more often to others' relationship with personal technology. Moreover, adolescents did not associate "Hurtful," "Overwhelming," or "Emotional" with their own or others' relationship with technology.

Adolescents associated "Provide an escape" more with their own technology, whereas they associated "Creates barrier" and "It's just a tool" more for describing others' relationship with personal technology. This difference has not been highlighted in prior research.



XSL•FO
RenderX

**Figure 2.** Cross-comparison between adolescents' perceptions of their own versus others' relationship with personal technology.




We were able to produce these findings by applying a multi-lens approach. A deeper understanding of the factors underlying these findings is critical and may have direct implications on technology designed for adolescents.

### *Understanding Adolescents' Perception of Their Current Relationship With Technology Versus Their Aspirations*

As mentioned earlier, adolescents selected "My personal assistant" the most when asked about their aspirations of a persona for their personal technology in question 3. We further cross-compared adolescents' aspirations (lens 3) and perceptions of current relationship with technology (lens 1). One interesting association which surfaced was that the adolescents who chose the "My best friend" persona for their personal technology, selected "Provides an escape'" as their top descriptor for their self-relationship with technology (27/148, 18.2% chose this option). Wishing for personal technology to be "My best friend" and seeing technology as something which "Provides an escape" shows critical associations that need further in-depth explorations.

In summary, our study indicates that adolescents see both commonalities and variations in their own and others' relationship with technology. Their future aspirations for personal technology vary across age and gender. We will validate our preliminary findings in a follow-up study using a larger sample size.

## Conclusions

This formative research explores adolescents' perception of relationship with personal technology. Unlike prior studies, our novel 3-lens approach is not limited to characterizing only adolescents' perception of technology as users. Instead, it further allows a comparison of adolescents' perception of technology as users versus bystanders and from current versus aspirational perspectives. The 3-lens approach yielded findings that show both alignment and conflict in perception of self-use versus others' use of personal technology. Our study also demonstrated variation in the perception of the youngest adolescents compared to the rest of the group as to how personal their relationship is with their personal technology.

Foundational understanding of adolescents' relationship perception through our multi-lens approach also offers a guiding perspective to personal technology user-experience designers. This research will also empower health care professionals and youth counselors to understand aligned and conflicted perceptions and design appropriate intervention for addressing the negative implications of technology.

In the follow-up research, we will conduct a series of focus groups with adolescents of different ages. We will focus on understanding the rationale behind adolescents' perception of the technology relationships and validate the findings from this study.

## Conflicts of Interest

None declared.

## References

1. Schaeffer K. Most US teens who use cellphones do it to pass time, connect with others, learn new things. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2019/08/23/most-u-s-teens-who-use-cellphones-do-it-to-pass-time-connect-with-others-learn-new-things/ [accessed 2020-02-20]
2. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. J Affect Disord 2017 Jan 01;207:251-259. [doi: 10.1016/j.jad.2016.08.030] [Medline: 27736736]



XSL•FO
RenderX

3. Victoria R, Susannah F, Well Being Trust. Digital health practices, social media use, and mental well-being among teens and young adults in the U.S. Providence St. Joseph Health Digital Commons. 2018. URL: https://digitalcommons.psjhealth.org/cgi/viewcontent.cgi?article=2092&context=publications [accessed 2021-09-02]
4. Uhls Y, Ellison N, Subrahmanyam K. Benefits and Costs of Social Media in Adolescence. Pediatrics 2017 Nov;140(Suppl 2):S67-S70 [FREE Full text] [doi: 10.1542/peds.2016-1758E] [Medline: 29093035]
5. Chen B, Liu F, Ding S, Ying X, Wang L, Wen Y. Gender differences in factors associated with smartphone addiction: a cross-sectional study among medical college students. BMC Psychiatry 2017 Oct 10;17(1):341 [FREE Full text] [doi: 10.1186/s12888-017-1503-z] [Medline: 29017482]
6. Vossen H, Valkenburg P. Do social media foster or curtail adolescents' empathy? A longitudinal study. Computers in Human Behavior 2016 Oct;63:118-124 [FREE Full text] [doi: 10.1016/j.chb.2016.05.040]
7. De-Sola Gutiérrez J, Rodríguez de Fonseca F, Rubio G. Cell-phone addiction: a review. Front Psychiatry 2016;7:175 [FREE Full text] [doi: 10.3389/fpsyt.2016.00175] [Medline: 27822187]
8. Jin B, Park N. In-person contact begets calling and texting: interpersonal motives for cell phone use, face-to-face interaction, and loneliness. Cyberpsychol Behav Soc Netw 2010 Dec;13(6):611-618 [FREE Full text] [doi: 10.1089/cyber.2009.0314] [Medline: 21142985]
9. Scott D, Valley B, Simecka B. Mental health concerns in the digital age. Int J Ment Health Addiction 2016 Jun 17;15(3):604-613 [FREE Full text] [doi: 10.1007/s11469-016-9684-0]
10. Yarosh S, Mcroberts S, Bonsignore E, Peyton T. YouthTube: Youth video authorship on youtube and vine. 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing 2016:1423-1437.
11. Rachel M, Denise A, Andrea F. Four factors that regulate teen technology use in everyday life. 2017 Presented at: The ACM Conference on Computer Supported Cooperative WorkSocial Computing (CSCW '17); Feb 26-March 12017; New York, USA p. 511-522 URL: https://doi.org/10.1145/2998181.2998310 [doi: https://doi.org/10.1145/2998181.2998310]
12. Danah B. It's Complicated The Social Lives of Networked Teens. New Haven: Yale University Press; Feb 24, 2015.
13. Amanda L. Pew Research Center Report. URL: http://www.pewinternet.org/2015/04/09/teens-social [accessed 2021-05-25]
14. Davis K, Dinhopl A. "Everything's the Phone": understanding the phone's supercharged role in parent-teen relationships. 2019 May 02 Presented at: The CHI conference on human factors in computing systems; May 4-9, 2019; Glasgow p. 1-14 URL: https://doi.org/10.1145/3290605.3300457 [doi: 10.1145/3290605.3300457]
15. Lopez-Fernandez O, Honrubia-Serrano L, Freixa-Blanxart M, Gibson W. Prevalence of problematic mobile phone use in British adolescents. Cyberpsychol Behav Soc Netw 2014 Mar;17(2):91-98. [doi: 10.1089/cyber.2012.0260] [Medline: 23981147]
16. Anderson M, Jiang J. Teens? Social media habits and experiences. Pew Research Center: Internet, Science & Tech. URL: https://www.pewresearch.org/internet/2018/11/28/teens-social-media-habits-and-experiences/ [accessed 2021-06-01]
17. Council on Communications and Media. Media se in school-aged children and adolescents. Pediatrics 2016 Nov 21;138(5):e20162592 [FREE Full text] [doi: 10.1542/peds.2016-2592] [Medline: 27940794]
18. Peck J, Stanton M, Reynolds G. Smartphone preventive health care: parental use of an immunization reminder system. Journal of Pediatric Health Care 2014 Jan;28(1):35-42. [doi: 10.1016/j.pedhc.2012.09.005] [Medline: 23195652]
19. Shin D, Shin Y, Choo H, Beom K. Smartphones as smart pedagogical tools: Implications for smartphones as u-learning devices. Computers in Human Behavior 2011 Nov;27(6):2207-2214 [FREE Full text] [doi: 10.1016/j.chb.2011.06.017]
20. Rung A, Warnke F, Mattheos N. Investigating the use of smartphones for learning purposes by Australian dental students. JMIR Mhealth Uhealth 2014 Apr 30;2(2):e20 [FREE Full text] [doi: 10.2196/mhealth.3120] [Medline: 25099261]
21. Cochrane T. Beyond the Yellow Brick Road: Mobile Web 2.0 informing a new institutional e-learning strategy. ALT-J 2016 Dec 13;18(3):221-231 [FREE Full text] [doi: 10.1080/09687769.2010.529110]
22. Burnett C, Merchant G. Points of view: reconceptualising literacies through an exploration of adult and child interactions in a virtual world. Journal of Research in Reading 2013 May 27;37(1):36-50. [doi: 10.1111/jrir.12006]
23. Thomée S. Mobile phone use and mental health. A review of the research that takes a psychological perspective on exposure 2018:1-25 [FREE Full text] [doi: 10.3390/ijerph15122692]
24. Hale L, Li X, Hartstein L, LeBourgeois M. Media use and sleep in teenagers: what do we know? Curr Sleep Medicine Rep 2019 Jul 16;5(3):128-134 [FREE Full text] [doi: 10.1007/s40675-019-00146-x]
25. Park N, Lee H. Nature of youth smartphone addiction in Korea: diverse dimensions of smartphone use and individual traits. journalofcommunicationresearch 2014 Feb;51(1):100-132 [FREE Full text] [doi: 10.22174/jcr.2014.51.1.100]
26. Kwon M, Kim D, Cho H, Yang S. The smartphone addiction scale: development and validation of a short version for adolescents. PLoS One 2013;8(12):e83558 [FREE Full text] [doi: 10.1371/journal.pone.0083558] [Medline: 24391787]
27. Hale L, Li X, Hartstein L, LeBourgeois M. Media use and sleep in teenagers: what do we know? Current Sleep Medicine Reports 2019 Jul 16;5(3):128-134 [FREE Full text] [doi: 10.1007/s40675-019-00146-x]
28. Aziz N, Nordin M, Abdulkadir S, Salih M. Digital addiction: systematic review of computer game addiction impact on adolescent physical health. Electronics 2021 Apr 22;10(9):996 [FREE Full text] [doi: 10.3390/electronics10090996]
29. Parasuraman S, Sam A, Yee SK, Chuon BC, Ren L. Smartphone usage and increased risk of mobile phone addiction: A concurrent study. Int J Pharm Investig 2017;7(3):125-131 [FREE Full text] [doi: 10.4103/jphi.JPHI_56_17] [Medline: 29184824]




XSL•FO
RenderX

30. Choi H, Lee H, Ha J. The influence of smartphone addiction on mental health, campus life and personal relations - Focusing on K university students. Article in Korean. Journal of the Korean Data and Information Science Society 2012 Sep 30;23(5):1005-1015. [doi: 10.7465/jkdi.2012.23.5.1005]
31. Nowland R, Necka E, Cacioppo T. Loneliness and social internet use: pathways to reconnection in a digital world? Perspect Psychol Sci 2018 Jan;13(1):70-87. [doi: 10.1177/1745691617713052] [Medline: 28937910]
32. Atarodi A, Rajabi M, Atarodi A. Cell phone use and social alienation of young teenagers. Library Philosophy and Practice 2020:1-10.
33. Jeong S, Kim H, Yum J, Hwang Y. What type of content are smartphone users addicted to?: SNS vs. games. Computers in Human Behavior 2016 Jan;54:10-17. [doi: 10.1016/j.chb.2015.07.035]
34. Kim H, Min J, Min K, Lee T, Yoo S. Relationship among family environment, self-control, friendship quality, and adolescents' smartphone addiction in South Korea: Findings from nationwide data. PLoS One 2018 Feb 5;13(2):e0190896 [FREE Full text] [doi: 10.1371/journal.pone.0190896] [Medline: 29401496]
35. Fırat S, Gül H, Sertçelik M, Gül A, Gürel Y, Kılıç B. The relationship between problematic smartphone use and psychiatric symptoms among adolescents who applied to psychiatry clinics. Psychiatry Res 2018 Dec;270:97-103 [FREE Full text] [doi: 10.1016/j.psychres.2018.09.015] [Medline: 30245382]
36. Cocoradă E, Maican CI, Cazan A, Maican MA. Assessing the smartphone addiction risk and its associations with personality traits among adolescents. Children and Youth Services Review 2018 Oct;93:345-354. [doi: 10.1016/j.childyouth.2018.08.006]
37. Yang Y, Yen J, Ko C, Cheng C, Yen C. The association between problematic cellular phone use and risky behaviors and low self-esteem among Taiwanese adolescents. BMC Public Health 2010 Apr 28;10(1):217 [FREE Full text] [doi: 10.1186/1471-2458-10-217] [Medline: 20426807]
38. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. Journal of Behavioral Addictions 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]
39. Anderson M, Jiang J. Teens, social media and technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2021-05-25]
40. Ozkan M, Solmaz B. Mobile addiction of generation z and its effects on their social lifes:(An application among university students in the 18-23 age group). Procedia-Social and Behavioral Sciences, 205, 2015 May:92-98 [FREE Full text] [doi: 10.1016/j.sbspro.2015.09.027]
41. Schoeni A, Roser K, Röösli M. Symptoms and cognitive functions in adolescents in relation to mobile phone use during night. PLoS One 2015;10(7):e0133528 [FREE Full text] [doi: 10.1371/journal.pone.0133528] [Medline: 26222312]
42. Activity vector analysis. AVA Assessment. URL: https://www.ava-assessment.com/ [accessed 2021-05-27]
43. Adjective Checklist (ACL). Statistics Solutions. 2018. URL: https://www.statisticssolutions.com/adjective-checklist-acl/ [accessed 2021-05-25]

*Edited by G Eysenbach; submitted 09.02.21; peer-reviewed by R Pine, A Jolliff, N Griffioen; comments to author 24.02.21; revised version received 21.07.21; accepted 27.07.21; published 23.12.21.*

*Please cite as:*
*Jee M, Khan A, Nazneen N*
*Understanding Adolescents' Perceptions and Aspirations Towards Their Relationship With Personal Technology: Survey Study*
*JMIR Form Res 2021;5(12):e27852*
*URL: https://formative.jmir.org/2021/12/e27852*
*doi:10.2196/27852*
*PMID:34941556*

©Muhammad Jee, Alsa Khan, Nazneen Nazneen. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# A Digital Health Fall Prevention Program for Older Adults: Feasibility Study

Claire L Jacobson[1,2], BA, MS; Lauren C Foster[1], BA; Hari Arul[1], BS; Amanda Rees[1], BS; Randall S Stafford[1,2], BA, MHS, MS, MD, PhD

[1]Age Bold Inc, Los Angeles, CA, United States

[2]Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, CA, United States

**Corresponding Author:**
Lauren C Foster, BA
Age Bold Inc
Los Angeles, CA
United States
Phone: 1 6503811395
Email: lauren@agebold.com

## Abstract

**Background:** About 1 in 3 adults aged 65 and older falls annually. Exercise interventions are effective in reducing the fall risk and fall rate among older adults. In 2020, startup company Age Bold Inc. disseminated the Bold Fall Prevention Program, aiming to reduce falls among older adults through a remotely delivered, digital exercise program.

**Objective:** We conducted a feasibility study to assess the delivery of the Bold Fall Prevention Program remotely and evaluate the program's impact on 2 primary outcomes—annualized fall rate and weekly minutes of physical activity (PA)—over 6 months of follow-up.

**Methods:** Older adults at high risk of falling were screened and recruited for the feasibility study via nationwide digital advertising strategies. Self-reported outcomes were collected via surveys administered at the time of enrollment and after 3 and 6 months. Responses were used to calculate changes in the annualized fall rate and minutes of PA per week.

**Results:** The remote delivery of a progressive digital fall prevention program and associated research study, including remote recruitment, enrollment, and data collection, was deemed feasible. Participants successfully engaged at home with on-demand video exercise classes, self-assessments, and online surveys. We enrolled 65 participants, of whom 48 (74%) were women, and the average participant age was 72.6 years. Of the 65 participants, 54 (83%) took at least 1 exercise class, 40 (62%) responded to at least 1 follow-up survey at either 3 or 6 months, 20 (31%) responded to both follow-up surveys, and 25 (39%) were lost to follow-up. Among all participants who completed at least 1 follow-up survey, weekly minutes of PA increased by 182% (ratio change=2.82, 95% CI 1.26-6.37, n=35) from baseline and annualized falls per year decreased by 46% (incidence rate ratio [IRR]=0.54, 95% CI 0.32-0.90, n=40). Among only 6-month survey responders (n=31, 48%), weekly minutes of PA increased by 206% (ratio change=3.06, 95% CI 1.43-6.55) from baseline to 6 months (n=30, 46%) and the annualized fall rate decreased by 28% (IRR=0.72, 95% CI 0.42-1.23) from baseline to 6 months.

**Conclusions:** The Bold Fall Prevention Program provides a feasible strategy to increase PA and reduce the burden of falls among older adults.

***(JMIR Form Res 2021;5(12):e30558)*** doi:10.2196/30558

**KEYWORDS**

older adults; accidental falls; fall prevention; digital health; technology; exercise; longevity and healthy aging; program evaluation; aging; elderly; health strategy

## Introduction

Every year, one-third of community-dwelling older adults (adults aged 65 and older) experience a fall [1]. Falls, defined as “unexpected event[s] in which the participant comes to rest on the ground, floor, or lower level,” are responsible for a wide range of negative health outcomes [2]. Annually, more than 3 million older Americans are treated in emergency departments for fall injuries, and over 800,000 are hospitalized, often because of a head injury or hip, spine, or wrist fracture [3]. Falls are the



XSL•FO
RenderX

leading cause of injury-related deaths among older adults, and the age-adjusted fall death rate (64 deaths per 100,000 older adults) increased by 30% from 2009 to 2018 [4]. Additionally, the psychological impact of falling can cause older adults and their caregivers significant fear about the risk of falling again. This fear of falling can have an accumulating effect whereby the fear of falling causes individuals to limit their everyday physical activities, which in turn makes them weaker and more susceptible to future falls. In fact, studies have shown that falling once doubles the chances of falling again [5].

These negative physical and psychological impacts of falling are accompanied by a staggering price tag for the American health care system. According to the Centers for Disease Control and Prevention (CDC), the direct medical costs for falls in 2015 alone totaled more than $50 billion, and 75% of those costs were shouldered by Medicare and Medicaid [6]. On average, the direct medical cost of hospitalization for a fall-related injury is $30,000 and increases with age [7]. As the number of older adults in the United States continues to grow, so too will the projected fall-related costs.

Many falls, however, can be prevented. One of the most effective ways to reduce fall risk is through targeted exercise that improves an individual's strength, balance, and mobility [8]. Exercise-based programs, such as the Otago Exercise Program and Tai Ji Quan: Moving for Better Balance, have been shown to reduce falls by up to 35% and 55%, respectively [9,10]. Until recently, however, the vast majority of fall prevention programs were only offered in small, in-person classes hosted in local senior centers or gyms. Although this has been the standard dissemination method for decades, it comes with significant barriers to participation. Common barriers to in-person programs include a lack of programs in rural or underresourced communities; limited or no access to transportation; scheduling conflicts; cost of getting to and using facilities; interpersonal barriers, such as finding other participants' presence intimidating; and physical environmental barriers, such as bad weather, stairs, uneven ground, difficult parking, and more [11].

In the past decade, trends in the United States indicate substantial movement toward the use of digital health tools for older adults. Between 2013 and 2017, smartphone ownership by those aged 65 and older more than doubled from 18% to 42% [12]. Similarly, 67% of all US older adults report that they spend time online [12]. Leveraging technology to facilitate remote fall prevention exercise programs provides an exciting opportunity to make a meaningful impact in elder care by increasing access to training and resources and removing barriers that have limited the reach of previous efforts. To meet this growing need, Age Bold Inc. (Bold), a San Francisco–based digital health company, created the Bold Fall Prevention Program. Here, we describe the results and impacts of the program's feasibility study launched in January 2020, including an evaluation of the two primary outcomes—annualized fall rates and minutes of physical activity (PA)—at 6-month follow-up.

## Methods

### Purpose and Objectives

The feasibility study for the Bold Fall Prevention Program was designed to demonstrate the ability to remotely deliver an innovative digital exercise program. We assessed the program's ability to engage and retain participants and to remotely collect outcomes data. The study also examined the program's impact on fall rates and measures of physical function in older adults over 6 months, with continued follow-up to 12 months. The research study was approved by the WIRB-Copernicus Group® (Puyallup, WA, USA) institutional review board (#520190289).

### Intervention Approach

Bold's Fall Prevention Program is a 12-week digital program of progressive exercise routines aimed at increasing strength, mobility, and balance to reduce the risk of falls. All exercise sessions (classes) were provided online and on demand. Every session was guided by Bold instructors, which included kinesiologists, personal trainers, and community Tai Chi instructors. The program included exercises common to evidence-based programs, such as Stay Active and Independent for Life (SAIL), FallProof, Matter of Balance, and Tai Ji Quan: Moving for Better Balance.

Bold's classes were designed to progressively build physical strength, balance, and mobility and included static and dynamic movements that challenged a participant's center of gravity (eg, tandem walking). The addition of weight-shifting and overreaching exercises progressively increased the challenge of these exercises throughout the duration of the program. The classes included a variety of exercises for activation of the anterior and posterior of the lower leg and foot muscles (eg, plantar and dorsiflexion), for hip activation (eg, squats, lunges, lateral leg lifts, knee lifts, and hip extensions), and for postural control (eg, chest stretches, chin tucks, and rows). Hamstring curls, hip extensions, hip flexion, and leg extensions were incorporated to build lower body strength, specifically in the quadriceps, glutes, and hamstrings. Side bends, knee to elbow, trunk rotations, and static abdominal holds were added to build upper and lower back strength. The program classes also included instruction of 8 Tai Chi forms in seated or standing positions (eg, parting wild horse's mane, single whip, wave hands like clouds, repulse monkey, brush knee, grasp the peacock's tail, and fair lady works shuttles). The majority of Bold programming combined elements from evidence-based fall prevention programs, and additional exercises were designed and taught by a kinesiologist with more than a decade of experience working with older adults.

Class durations ranged from 30 to 45 minutes per class, 3 times a week, for a minimum of 100 minutes of class instruction per week. This exercise dosage was based on findings from early pilot testing and is in conjunction with other validated fall prevention programs [13,14]. Participants were encouraged to take a Bold class at least 2 times per week, supplemented by other light PA, like walking, on most other days.

Based on the participants' stated preferences and abilities, collected at baseline, they received a customized program that



XSL•FO
RenderX

began with either seated or standing exercises and progressed in difficulty over time. Ankle weights and a balance prop (a balance disc for participants beginning seated and a balance pad for participants beginning standing) were shipped to participants upon enrollment in the program. In the seated exercise progression, ankle weights were introduced for strength-focused sessions, a balance disc was used to build core strength and postural control in advanced seated balance exercises, and as the program progressed, participants gradually began incorporating standing exercises into their routines. In the standing progression, ankle weights were introduced for strength-focused sessions, and a foam balance pad was introduced for advanced standing balance exercises.

A typical class structure included a dynamic warm-up; a balance, strength, or combination balance-and-strength segment; and a cool-down (Multimedia Appendix 1). In each class, instructors provided safety guidelines and instructions for when to use props or modify various exercises. Participants were also encouraged to reach out to the instructors and research team at the end of each class with questions about using the props and to get further guidance, if so desired.

In addition to weekly exercise classes, participants received educational and motivational support, including weekly emails with resources to learn more about balance and fall prevention, individual motivational text messages, and phone calls. After the initial 12-week program was completed, participants progressed to a 9-month "maintenance" phase with access to additional exercise classes to maintain their progress.

## Recruitment

Potential participants were identified and recruited through nationwide digital strategies and physical flyers placed in local communities in California. Prospective participants were directed to learn more about the study on the study website.

Eligible participants were 65 years old or older, had reliable computer or tablet access with broadband internet, and were able to consent and follow study instructions in English. In addition, they met 1 or more of the following criteria: (1) a history of 1 or more injurious fall(s) in the past 12 months, (2) a history of 2 or more non-injurious falls in the past 12 months, (3) reduced mobility indicated by a Timed Up & Go test score of ≥12 seconds ("mobility assessment"), (4) reduced balance indicated by an inability to hold positions 3 or 4 for ≥10 seconds on the 4-Stage Balance Test ("balance assessment"), or (5) reduced strength indicated by a score below average for age/gender on the 30-second Chair Stand Test ("strength assessment").

The mobility, balance, and strength assessments were directly adapted from the CDC's Stopping Elderly Accidents, Deaths, and Injuries functional assessments [15]. These assessments were adapted to the Bold platform through several rounds of user testing prior to the start of the study. During these sessions, older adult participants would conduct the self-assessment while a member of the research team observed. Feedback gathered from participants and the research team during these sessions informed how the self-assessments were presented on the digital platform and the guidance provided to study participants. Delivered remotely via the Bold website during the eligibility screening, participants watched an instructional video in which a Bold instructor explained how to complete the self-assessment and demonstrated each step in the process. Participants were able to watch the video as often as they wanted before they began the assessment. For the strength and mobility self-assessments, participants were provided with a digital stopwatch that they could control, as needed. For the strength assessment, they were instructed to count the number of repetitions they could complete within the allotted time. For the mobility assessment, they could start and stop the stopwatch, as needed, and adjust their final time to accommodate any delays between completing the assessment and hitting the digital stopwatch.

Participants were excluded if they met any of the following criteria: (1) regular use of a mobility aid, such as a walker, wheelchair, or scooter; (2) any contraindication to light PA; (3) neurological impairments that impact gait (eg, Parkinson's disease); (4) cognitive impairment above a level consistent with safety; (5) severely reduced measures of physical function (eg, unable to complete functional assessments); and (6) at the clinical discretion of investigators.

In total, 65 participants who met the eligibility requirements were enrolled in the study. Each participant received $10 for completion of the baseline survey, $10 for completion of the 3-month survey, and $20 for completion of the 6-month survey.

## Data Collection and Statistical Analysis

Eligibility and outcome data were remotely collected via questionnaires on Bold's website during the enrollment process. Follow-up surveys were administered via Typeform, an online survey tool that allows the remote delivery, collection, and tracking of study surveys to participants.

To evaluate the feasibility of the intervention, we focused on 3 areas: (1) the feasibility of remote enrollment of a population at high risk of falls, (2) the feasibility of remote data collection, and (3) the feasibility of digitally delivering a progressive fall prevention exercise program.

To assess the program's impact on PA, participants were asked to think about the previous 7 days and report how many minutes of PA (including walking) they did each day. Responses were used to estimate the weekly minutes of PA at baseline, 3 months, 6 months, and a combined 3-and 6-month estimate (weighted by the person-years since the last survey). Weekly minutes of PA were logarithmically transformed to normalize the data and allow for statistical analysis. Unpaired and paired 2-tailed *t* tests were used to calculate mean group differences of the log-transformed minutes of PA. Exponentiating these differences provided estimates of the ratio change from baseline.

To evaluate the program's impact on fall rates and weekly minutes of PA, enrolled participants completed surveys at the time of enrollment (baseline survey), 3 months after enrollment, and 6 months after enrollment. In each survey, participants were asked to report the total number of falls experienced, if any, since the last survey. At 3 and 6 months, the annualized fall rate was calculated as the total number of falls reported divided by 0.25 and 0.5 person-years, respectively [1,14,16]. For



XSL•FO
RenderX

participants who completed both surveys, the total number of falls at 6 months was inclusive of any falls previously reported in the 3-month survey. Changes in annualized falls rates were calculated as incidence rate ratios (IRRs) from baseline to 3 months, 6 months, and combined 3 and 6 months. All change estimates are reported with 95% CIs and 2-tailed *P*-values. Descriptive statistical methods, including mean, standard deviations, and frequencies, were also used to summarize study data. Statistical analyses were performed using R Studio software (Boston, MA, USA).

## Results

### Demographic Characteristics of Participants

The demographic characteristics of the study population are presented below (Table 1). The study population included 65 participants: 48 (74%) women and 17 (26%) men. Of these, 60 (92%) participants identified as White, and the mean age was 72.6 years. The highest education level attained was high school or equivalent for 7 (11%) participants, some college or associate degree for 16 (25%) participants, and college graduate or higher for 42 (65%) participants. In addition, 58 (89%) participants had Medicare or Medicare Advantage as their health insurance. Furthermore, 30 (46%) participants had at least 1 chronic condition and 12 (19%) had at least 5 prescribed medications. The majority (39/65, 60%) of participants had fallen at least 1 time in the previous 12 months, and on average, participants engaged in 151 minutes of PA per week, although almost half (30/65, 46%) of the participants reported engaging in less than 1 hour of PA per week.



XSL•FO
RenderX

**Table 1.** Baseline characteristics of study participants (N=65).

| Characteristics | Number of participants, n (%) |
|---|---|
| **Gender** | |
| Female | 48 (74) |
| Male | 17 (26) |
| **Age in years** | |
| 65-69 | 25 (39) |
| 70-74 | 22 (34) |
| 75-79 | 8 (12) |
| 80 and older | 10 (15) |
| **Highest education level** | |
| High school diploma or equivalent | 7 (10) |
| Some college or associate degree | 16 (25) |
| College graduate or higher | 42 (65) |
| **Race/ethnicity** | |
| White | 60 (93) |
| American Indian or Alaska Native | 1 (1) |
| Asian (including South Asian and Asian Indian) | 1 (1) |
| Black or African American | 2 (4) |
| Prefer not to state | 1 (1) |
| **Income in US $** | |
| <20,000 | 9 (14) |
| 20,000-49,999 | 21 (32) |
| 50,000-74,999 | 18 (28) |
| 75,000-99,999 | 8 (12) |
| ≥100,000 | 5 (8) |
| Prefer not to state | 4 (6) |
| **Geographic region** | |
| West | 20 (30) |
| Midwest | 14 (21) |
| Northeast | 17 (26) |
| South | 14 (22) |
| **Insurance** | |
| Medicare | 40 (61) |
| Medicare Advantage Private Plan (Medicare Part C) | 18 (28) |
| Employer-based insurance | 5 (8) |
| Veterans Affairs Health Care | 2 (3) |
| **Chronic medical conditions**[a] | |
| Any chronic medical condition | 30 (46) |
| Cardiovascular disease | 9 (14) |
| Type 2 diabetes | 12 (19) |
| Musculoskeletal condition | 24 (37) |
| Chronic obstructive pulmonary disorder (COPD)/other lung disease | 5 (8) |
| Depression | 7 (11) |



XSL•FO
RenderX

| Characteristics | Number of participants, n (%) |
|---|---|
| Other | 12 (19) |
| **Number of prescribed medications** | |
| None | 15 (23) |
| 1-4 | 38 (59) |
| ≥5 | 12 (18) |
| **Fall in previous 12 months** | |
| Yes | 39 (60) |
| No | 26 (40) |
| **Amount of weekly PA[b] in minutes** | |
| <60 | 29 (45) |
| 61-120 | 9 (14) |
| 121-180 | 9 (14) |
| 181-240 | 4 (6) |
| 241-300 | 4 (6) |
| 201-360 | 1 (1) |
| >360 | 7 (11) |

[a]Multiple responses were possible.
[b]PA: physical activity.

## Feasibility

In all, 143 high-fall-risk older adults were recruited between January and February 2020. Tracking showed that 100% of the eventually enrolled participants were found via digital ads on Facebook. Of the 143 older adults recruited, 65 (45.5%) completed enrollment. Most of this drop-off can be attributed to individuals who began the enrollment process because they were curious, but ceased involvement after learning about the time commitment required for participation. The process of remote enrollment via online surveys and self-assessments was successful in identifying and engaging high-fall-risk older adults in the study. Participants did not report any issues with completing the surveys or administering the self-assessments.

Of the 65 enrolled participants, 54 (83%) took at least 1 class during the study period, 30 (46%) took at least 10 classes, and 14 (22) took at least 35 classes. The average number of classes taken was 18 (SD 21, median 7, 25th percentile=1, 75th percentile=33). In addition, 29 (45%) and 31 (48%) participants completed the 3- and 6-month surveys, respectively, indicating that the remote data collection process was only moderately successful in achieving high levels of survey completion. Operationally, automated data collection, tracking, and alert notifications were beneficial to the research team and greatly diminished the need for a large research staff. However, during the collection of the 3-month follow-up survey, the research team noticed that most participants required between 1 and 3 phone calls to prompt them to complete their survey, suggesting that establishing a connection with a member of the research team was an integral part of the data collection process. Accordingly, it was determined that although the process of remote data collection provides distinct benefits for a research team working with a geographically distributed population, this modality requires additional support to mitigate substantial losses to follow-up. These findings indicate that it is feasible for older adults to adhere to a digital exercise program; however, they would likely benefit from supplemental outreach to encourage consistent class-taking and engagement behavior.

From the perspective of both the participants and the research team, the digital delivery of a progressive fall prevention exercise program was feasible. Participants were able to engage in the program and follow along week by week, even during the COVID-19 pandemic. Many participants reported the program as their sole form of exercise due to the many of the pandemic-related restrictions put in place, and stated that they wished to continue exercising at home even after the restrictions were lifted. This suggests that the digital exercise program addresses a previously unmet need in the community and its delivery was sufficient to engage those community members. For the research team, in addition to strong relationships with the engineering team, which allowed for quick troubleshooting of technical issues, traditional methods, such as calling and emailing participants, to check in and establish connections were crucial. Doing so allowed the research team to answer participant questions quickly, anticipate future challenges, and proactively respond to aspects of the program experience that may be confusing. Overall, the remote delivery of the program was sufficient to establish a basis for the further development and expansion of the digital fall prevention exercise program.

## Physical Activity

The change in weekly minutes of physical activity for all participants in shown in Table 2. Paired comparisons showed that weekly minutes of PA were 3.06 times higher (95% CI



XSL•FO
RenderX

1.43-6.55) at 6 months than at baseline (n=30, 46%; Table 3). At 3 months (n=18, 28%), weekly minutes of PA were 1.49 times higher (95% CI 0.60-3.72) than at baseline (Table 3). The combined estimate of all participants who submitted at least 1 follow-up survey with completed PA data (n=35, 54%) showed that weekly minutes of PA were 2.82 times higher (95% CI 1.26-6.37) than at baseline (n=35, 54%), a 182% increase (Table 3).

**Table 2.** Change in weekly minutes of physical activity for all participants (N=65).

| Time of survey, n (%) | Mean PA[a] in minutes | Mean log PA in minutes | Log SD | Difference from baseline in log minutes, n (95% CI) | Ratio change in minutes, n (95% CI) | *P* value[b] |
|---|---|---|---|---|---|---|
| Baseline, 63 (97) | 151 | 1.86 | 0.61 | Reference group | Reference group | __[c] |
| 3 months, 18 (28) | 251 | 2.24 | 0.58 | 0.38 (0.055-0.69) | 2.37 (1.14-4.90) | .02 |
| 6 months, 30 (46) | 456 | 2.38 | 0.65 | 0.52 (0.23-0.80) | 3.31 (1.71-6.31) | <.001 |
| Combined responders, 35 (54) | 396 | 2.34 | 0.78 | 0.48 (0.17-0.78) | 3.02 (1.49-6.07) | .003 |

[a]PA: physical activity.

[b]*P* value from *t* test.

[c]Not applicable.

**Table 3.** Subcohort analysis: baseline weekly minutes of physical activity by survey response group (N=65).

| Time of survey, n (%) | Mean PA[a] in minutes | Mean log PA in minutes | Log SD | Difference from baseline in log minutes, n (95% CI) | Ratio change in minutes, n (95% CI) | *P* value[b] |
|---|---|---|---|---|---|---|
| Baseline: 3-month responders, 18 (28) | 192 | 2.06 | 0.57 | 0.17 (–0.22 to 0.57)[c] | 1.49 (0.60-3.72)[c] | .37 |
| Baseline: 6-month responders, 30 (46) | 175 | 1.89 | 0.67 | 0.49 (0.16-0.82)[d] | 3.06 (1.43-6.55)[d] | .01 |
| Baseline: combined responders, 35 (54) | 162 | 1.89 | 0.63 | 0.45 (0.10-0.80)[e] | 2.82 (1.26-6.37)[e] | .01 |

[a]PA: physical activity.

[b]*P* value from *t* test.

[c]Calculated as 3 months vs baseline: 3-month responders.

[d]Calculated as 6 months vs baseline: 6-month responders.

[e]Calculated as combined responders vs baseline: combined responders.

## Falls

Of the 65 enrolled participants, 40 (62%) responded to at least 1 follow-up survey at either 3 or 6 months, 20 (31%) responded to both follow-up surveys, and 25 (39%) were lost to any follow-up. The change in the annualized fall rate for all participants is shown in Table 4. Among 6-month survey responders (n=31, 48%), the annualized fall rate decreased by 28% (IRR=0.72, 95% CI 0.42-1.23) from baseline to 6 months (Table 5). Among 3-month survey responders (n=29, 45%), the annualized fall rate per person decreased by 83% (IRR=0.17, 95% CI 0.05-0.55) from baseline to 3 months (Table 5). The combined estimate of all participants who completed at least 1 follow-up survey (n=40, 62%) showed a 46% reduction in the annualized fall rate (IRR=0.54, 95% CI 0.32-0.90) from baseline (Table 5).

**Table 4.** Change in annualized fall rate for all participants (N=65).

| Time of survey, n (%) | Fall rate (falls per person-year), n | SD | IRR[a] (95% CI) | *P* value[b] |
|---|---|---|---|---|
| Baseline, 65 (100) | 1.42 | 1.98 | Reference group | __[c] |
| 3 months, 29 (45) | 0.41 | 1.64 | 0.29 (0.09-0.92) | .03 |
| 6 months, 31 (48) | 1.16 | 3.09 | 0.82 (0.50-1.36) | .44 |
| Combined responders, 40 (62) | 1.01 | 1.17 | 0.72 (0.43-1.19) | .19 |

[a]IRR: incidence rate ratio.

[b]*P* value from the IRR.

[c]Not applicable.



XSL•FO
RenderX

**Table 5.** Subcohort analysis: baseline fall rate by survey response group (N=65).

| Time of survey, n (%) | Fall rate (falls per person-year), n | SD | IRR[a] (95% CI) | *P* value[b] |
|---|---|---|---|---|
| Baseline: 3-month responders, 29 (45) | 2.38 | 2.53 | 0.17 (0.05-0.55)[c] | <.001 |
| Baseline: 6-month responders, 31 (48) | 1.61 | 2.16 | 0.72 (0.42-1.23)[d] | .23 |
| Baseline: combined responders, 40 (62) | 1.88 | 2.32 | 0.54 (0.32-0.90)[e] | .02 |

[a]IRR: incidence rate ratio.

[b]*P* value from the IRR.

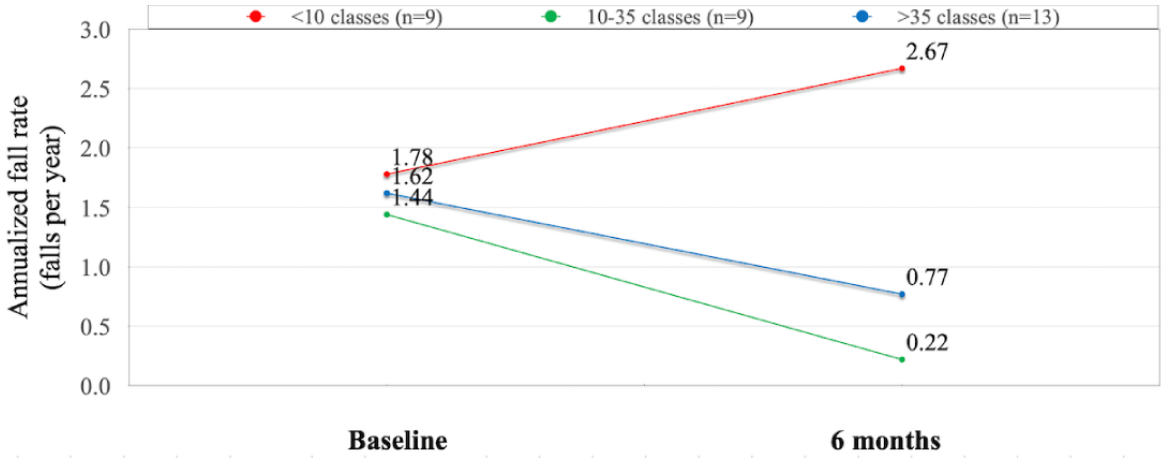
[c]Calculated as 3 months vs baseline: 3-month responders.

[d]Calculated as 6 months vs baseline: 6-month responders.

[e]Calculated as combined responders vs baseline: combined responders.

The change in the annualized fall rate correlated with the number of Bold classes taken among the 6-month survey responders. Participants who took fewer than 10 classes (n=9, 29%) reported an increase in the annualized fall rate from 1.78 annualized falls per person at baseline to 2.67 at 6 months, while those who took 10-35 classes (n=9, 29%) showed a decreased fall rate from 1.44 at baseline to 0.22 at 6 months, and those who took more than 35 classes (n=13, 42%) also showed a decrease from 1.62 at baseline to 0.77 at 6 months (Figure 1).

**Figure 1.** Change in the annualized fall rate per person between baseline and 6 months, stratified by class-taking behavior, for 31 of 65 participants (48%) who completed their 6-month survey.




There were some differences in baseline demographics between the groups. Of those who took more than 35 classes (n=13, 42%), 4 (31%) were over the age of 80 as compared to 1 (11%) in each of the other 2 cohorts (n=9 each). Of those who took fewer than 10 classes (n=9, 29%), 5 (56%) had experienced 1 or more falls in the 12 months prior to joining the study compared to 6 (67%) and 8 (62%) of those who took 10-35 (n=9, 29%) and more than 35 (n=13, 42%) classes, respectively. This cohort also had lower levels of average baseline PA as compared to those who took 10–35 or more than 35 classes, engaging in 116 minutes per week compared to 191 and 207 minutes, respectively. Further comparisons of demographic characteristics by class-taking behavior can be found in Multimedia Appendix 2.

Significant loss to follow-up was observed at 3 and 6 months. Baseline fall rates and PA levels, however, were similar for 3- and 6-month responders compared to all participants. This suggests that follow-up was not biased toward participants with higher baseline levels of PA.

## Discussion

### Principal Findings

The feasibility study findings indicate that it is possible to digitally disseminate this exercise-based fall prevention intervention and remotely study its impact on outcomes. At present, the majority of fall prevention programs occur in-person at local community centers, hospitals, or gyms, so the potential to deliver customized, digital-first programming could significantly increase the accessibility of such programs to people who cannot use the in-person offerings. It is also significant to the continued development of future studies of this intervention and provides useful guidance for future study design. From an effectiveness perspective, the study suggests that the Bold Fall Prevention Program has a positive impact on fall- and PA-related outcomes that endures even after the initial 3-month intervention window, with a notable 206% increase in PA from baseline to 6 months. The remote implementation and evaluation of a digital fall prevention program among older adults has the potential for significant health and cost impact among the rapidly growing older adult population in the United States.



XSL•FO
RenderX

## Comparison With Prior Work

Compared with in-person interventions, the Bold program reduced barriers to participant engagement by eliminating transportation challenges, allowing for self-scheduling of classes, customizing class content to match individuals' needs, and removing the need to have a local program facilitator or fitness center. An entirely digital program has the added benefits of scalability and lower distribution costs compared with traditional fall prevention programs.

Notably, the feasibility study was initiated just before the COVID-19 pandemic emerged in early 2020. Hence, participants likely were impacted by the countrywide trends of increased time spent at home, decreased PA, and increased feelings of isolation and stress, particularly for older adults who are most at risk for developing severe COVID-19 symptoms [17]. At the same time, COVID-19 provided a unique opportunity to disseminate and evaluate a digital health tool. The reliance on and growth of the remote delivery of health care and exercise will likely persist in the postpandemic world [18].

Among participants who responded to at least 1 follow-up survey (n=40, 62%), the statistically significant 46% reduction in the annualized fall rate (IRR=0.54) is slightly greater than that reported by the US Preventive Services Task Force (USPSTF) synthesis. The USPSTF data synthesis indicated that multifactorial PA interventions based on initial fall risk assessments are, on average, associated with a 21% reduction in the fall rate (IRR=0.79, 95% CI 0.68-0.91) [19]. Additionally, the observed 46% reduction in the fall rate during follow-up is comparable to the fall reductions reported by the Otago Exercise Program (reduced falls by 35%) and Tai Ji Quan: Moving for Better Balance (reduced falls by 55%), which are 2 of the most widely disseminated fall prevention programs in the United States [9,10].

## Limitations

There were several limitations of the feasibility study, including the number of participants lost to follow-up and the resultantly small sample size, reliance on self-reported measures for falls and PA, inconsistent survey response rates, and the lack of a control population, instead relying on before-after assessment of outcomes. It is hypothesized that the high loss to follow-up seen in the follow-up surveys was the result of a mix of contributing factors, including (1) user experience of difficulty when participants were directed to a website external to the Bold web application for the completion of their follow-up surveys, (2) a loss of the sense of personal connection and contribution to science as a result of the fully remote and digital nature of the study and study setting, and (3) the natural attrition that occurs in a study population over time. This is particularly informative as a lack of engagement and loss to follow-up are a particular concern for remote studies. Although the data presented in this paper are preliminary and best suited to be used as an early foundation for more formal investigations, the findings about the role that interpersonal connection, even over the phone or via email, can play in mitigating this loss to follow-up are important and should be strongly considered when planning future studies of this intervention.

The feasibility study was designed to demonstrate the ability to remotely deliver an innovative digital exercise program and provide preliminary evidence about the effectiveness of that program while also not obstructing the fast-paced and iterative nature of a startup. As such, Bold decided to conduct this feasibility study rather than immediately launching a resource-intensive and expensive randomized controlled trial. Although the research pursued is still valuable and unique within the for-profit digital health landscape, the findings are subject to more bias, confounding, and other systematic errors than would those of a randomized controlled trial.

The study did not define a clear minimum number of Bold classes that participants needed to complete to be included in the analysis. It is thought that 29 classes constitute a "full dose" of programming based on the number of minutes of strength, balance, and mobility training delivered per week, as well as the total number of targeted training sessions someone would typically receive after 29 classes (~18 hours). Although this is in line with other programs, such as SAIL and Matter of Balance, further research and investigation into this threshold is warranted. Additionally, we considered that enrolling into the program demonstrated an intention to reduce one's risk of falling and thus is an important early indicator of a subsequent reduction in falls. Accordingly, further investigation is warranted into the effect of class-taking behavior on the effectiveness of the intervention. These limitations notwithstanding, this feasibility study is valuable for assessing the basic feasibility and practicality of the intervention as well as for identifying areas for product improvements and modifications.

## Future Directions

Bold has used these study findings and participants' feedback to improve its current, publicly available product. Based on this feasibility study, we plan to conduct a clinical trial that rigorously assesses the impact of the program on fall rates, physical function, and health care costs in a Medicare Advantage population. Bold as a company is growing and widening its aperture to expand its video and nonvideo content to support more lasting behavior change and provide a more comprehensive toolkit to reduce older adults' likelihood of falling and improve their healthspan.

## Conclusion

Falls pose a significant public health threat for the increasing older adult population. Multifactorial and exercise-based fall prevention programs can reduce the risk of falls, yet the current landscape of fall prevention programs largely consists of fragmented, in-person, resource-intensive programs. The movement toward digital health tools, accelerated by the COVID-19 pandemic, highlights the timeliness and potential value of a digital exercise program designed specifically for older adults [12,15]. The Bold Fall Prevention Program leverages technology to provide an individually tailored program of online and on-demand exercise classes aimed at preventing falls. Preliminary results from this feasibility study suggest that the Bold Fall Prevention Program, with its potential for expanding access, could be a useful tool in reducing falls and increasing PA in older adults.



XSL•FO
RenderX

## Acknowledgments

Financial support for this work was provided by Age Bold Inc.

## Conflicts of Interest

All authors were employees of or advisors to Age Bold Inc at the time of research and analysis. Dr Stafford's contribution to this publication was as a consultant to Age Bold Inc. Dr Stafford is not providing this material as part of his Stanford University duties or responsibilities.

Multimedia Appendix 1
Exercise program progressions and sample class plans.
[DOCX File , 14 KB - formative_v5i12e30558_app1.docx ]

Multimedia Appendix 2
Participant demographics by class-taking behavior.
[DOCX File , 18 KB - formative_v5i12e30558_app2.docx ]

## References

1. Sherrington C, Fairhall NJ, Wallbank GK, Tiedemann A, Michaleff ZA, Howard K, et al. Exercise for preventing falls in older people living in the community. Cochrane Database Syst Rev 2019 Jan 31;1:CD012424 [FREE Full text] [doi: 10.1002/14651858.CD012424.pub2] [Medline: 30703272]
2. Carande-Kulis V, Stevens JA, Florence CS, Beattie BL, Arias I. A cost-benefit analysis of three older adult fall prevention interventions. J Safety Res 2015 Feb;52:65-70 [FREE Full text] [doi: 10.1016/j.jsr.2014.12.007] [Medline: 25662884]
3. Centers for Disease Control and Prevention. Important Facts about Falls. URL: https://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html [accessed 2021-04-29]
4. Centers for Disease Control and Prevention. Older Adult Falls Reported by State. URL: https://www.cdc.gov/homeandrecreationalsafety/falls/data/falls-by-state.html [accessed 2021-04-29]
5. O'Loughlin JL, Robitaille Y, Boivin JF, Suissa S. Incidence of and risk factors for falls and injurious falls among the community-dwelling elderly. Am J Epidemiol 1993 Feb 01;137(3):342-354. [doi: 10.1093/oxfordjournals.aje.a116681] [Medline: 8452142]
6. Florence CS, Bergen G, Atherly A, Burns E, Stevens J, Drake C. Medical costs of fatal and nonfatal falls in older adults. J Am Geriatr Soc 2018 Apr;66(4):693-698 [FREE Full text] [doi: 10.1111/jgs.15304] [Medline: 29512120]
7. Burns ER, Stevens JA, Lee R. The direct costs of fatal and non-fatal falls among older adults: United States. J Safety Res 2016 Sep;58:99-103 [FREE Full text] [doi: 10.1016/j.jsr.2016.05.001] [Medline: 27620939]
8. Stevens J, Burns E. A CDC Compendium of Effective Fall Interventions: What Works for Community-Dwelling Older Adults, 3rd ed. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2015.
9. Robertson MC, Campbell AJ, Gardner MM, Devlin N. Preventing injuries in older people by preventing falls: a meta-analysis of individual-level data. J Am Geriatr Soc 2002 May;50(5):905-911. [doi: 10.1046/j.1532-5415.2002.50218.x] [Medline: 12028179]
10. Li F, Harmer P, Fisher KJ, McAuley E, Chaumeton N, Eckstrom E, et al. Tai Chi and fall reductions in older adults: a randomized controlled trial. J Gerontol A Biol Sci Med Sci 2005 Feb;60(2):187-194. [doi: 10.1093/gerona/60.2.187] [Medline: 15814861]
11. Bethancourt HJ, Rosenberg DE, Beatty T, Arterburn DE. Barriers to and facilitators of physical activity program use among older adults. Clin Med Res 2014 Sep;12(1-2):10-20 [FREE Full text] [doi: 10.3121/cmr.2013.1171] [Medline: 24415748]
12. Anderson M, Perrin A. Technology Use Among Seniors.: Pew Research Center; 2019. URL: https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/ [accessed 2021-12-18]
13. The National Council on Aging. Evidence-Based Program: A Matter of Balance. URL: https://www.ncoa.org/article/evidence-based-program-a-matter-of-balance [accessed 2021-04-29]
14. Sherrington C, Michaleff ZA, Fairhall N, Paul SS, Tiedemann A, Whitney J, et al. Exercise to prevent falls in older adults: an updated systematic review and meta-analysis. Br J Sports Med 2017 Dec;51(24):1750-1758. [doi: 10.1136/bjsports-2016-096547] [Medline: 27707740]
15. Centers for Disease Control and Prevention. Clinical Resources: Functional Assessments. URL: https://www.cdc.gov/steadi/materials.html [accessed 2021-04-29]
16. Li F, Harmer P, Eckstrom E, Fitzgerald K, Chou L, Liu Y. Effectiveness of Tai Ji Quan vs multimodal and stretching exercise interventions for reducing injurious falls in older adults at high risk of falling: follow-up analysis of a randomized



XSL•FO
RenderX

clinical trial. JAMA Netw Open 2019 Feb 01;2(2):e188280 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.8280] [Medline: 30768195]
17. Mills JP, Kaye KS, Mody L. COVID-19 in older adults: clinical, psychosocial, and public health considerations. JCI Insight 2020 May 21;5(10):e139292 [FREE Full text] [doi: 10.1172/jci.insight.139292] [Medline: 32302293]
18. Alexander GC, Tajanlangit M, Heyward J, Mansour O, Qato DM, Stafford RS. Use and content of primary care office-based vs telemedicine care visits during the COVID-19 pandemic in the US. JAMA Netw Open 2020 Oct 01;3(10):e2021476 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21476] [Medline: 33006622]
19. Guirguis-Blake JM, Michael YL, Perdue LA, Coppola EL, Beil TL. Interventions to prevent falls in older adults: updated evidence report and systematic review for the US preventive services task force. JAMA 2018 Apr 24;319(16):1705-1716. [doi: 10.1001/jama.2017.21962] [Medline: 29710140]

## Abbreviations

**CDC:** Centers for Disease Control and Prevention
**IRR:** incidence rate ratio
**PA:** physical activity
**SAIL:** Stay Active and Independent for Life
**USPSTF:** US Preventive Services Task Force

*Edited by G Eysenbach; submitted 19.05.21; peer-reviewed by T Shubert, MN Shalaby; comments to author 21.07.21; revised version received 30.08.21; accepted 27.11.21; published 23.12.21.*

*Please cite as:*
*Jacobson CL, Foster LC, Arul H, Rees A, Stafford RS*
*A Digital Health Fall Prevention Program for Older Adults: Feasibility Study*
*JMIR Form Res 2021;5(12):e30558*
*URL: https://formative.jmir.org/2021/12/e30558*
*doi:10.2196/30558*
*PMID:34837492*

©Claire L Jacobson, Lauren C Foster, Hari Arul, Amanda Rees, Randall S Stafford. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Development of a Credible Virtual Clinician Promoting Colorectal Cancer Screening via Telehealth Apps for and by Black Men: Qualitative Study

Danyell Wilson-Howard[1,2], BSc, MA, PhD; Melissa J Vilaro[2], BA, MPH, CPH, PhD; Jordan M Neil[3,4], PhD; Eric J Cooks[2,5], BA, PhD; Lauren N Griffin[2], PhD; Taylor T Ashley[1], BS; Fatemeh Tavassoli[6], BA, PhD; Mohan S Zalake[6], BA; Benjamin C Lok[6], BSc, PhD; Folakemi T Odedina[7], PhD; Francois Modave[8], PhD; Peter J Carek[9], MD, MSc, CAQSM, FAAFP, DABF; Thomas J George[10,11], MD, FACP; Janice L Krieger[2,5], BSc, PhD

[1]Department of Natural Sciences, Bethune Cookman University, Daytona, FL, United States
[2]STEM Translational Communication Center, College of Journalism, University of Florida, Gainesville, FL, United States
[3]Harvard Medical School, Harvard University, Boston, MA, United States
[4]Mongan Institute's Health Policy Research Center, Massachusetts General Hospital, Boston, MA, United States
[5]UF Health Cancer Center, University of Florida, Gainesville, FL, United States
[6]Computer and Information Science and Engineering, University of Florida, Gainesville, FL, United States
[7]Center for Health Equity & Community Engagement Research, Mayo Clinic, Jacksonville, FL, United States
[8]Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, United States
[9]Family Medicine, Medical University of South Carolina, Charleston, SC, United States
[10]Gastrointestinal (GI) Oncology Center, University of Florida, Gainesville, FL, United States
[11]UF Health Medical Oncology – Davis Cancer, University of Florida, Gainesville, FL, United States

**Corresponding Author:**
Danyell Wilson-Howard, BSc, MA, PhD
Department of Natural Sciences
Bethune Cookman University
640 Mary McLeod Blvd.
Daytona, FL, 32114
United States
Phone: 1 3864812687
Email: wilsonda@cookman.edu

## Abstract

**Background:** Traditionally, promotion of colorectal cancer (CRC) screening among Black men was delivered by community health workers, patient navigators, and decision aids (printed text or video media) at clinics and in the community setting. A novel approach to increase CRC screening of Black men includes developing and utilizing a patient-centered, tailored message delivered via virtual human technology in the privacy of one's home.

**Objective:** The objective of this study was to incorporate the perceptions of Black men in the development of a virtual clinician (VC) designed to deliver precision messages promoting the fecal immunochemical test (FIT) kit for CRC screening among Black men in a future clinical trial.

**Methods:** Focus groups of Black men were recruited to understand their perceptions of a Black male VC. Specifically, these men identified source characteristics that would enhance the credibility of the VC. The modality, agency, interactivity, and navigability (MAIN) model, which examines how interface features affect the user's psychology through four affordances (modality, agency, interactivity, and navigability), was used to assess the presumed credibility of the VC and likability of the app from the focus group transcripts. Each affordance triggers heuristic cues that stimulate a positive or a negative perception of trustworthiness, believability, and understandability, thereby increasing source credibility.

**Results:** In total, 25 Black men were recruited from the community and contributed to the development of 3 iterations of a Black male VC over an 18-month time span. Feedback from the men enhanced the visual appearance of the VC, including its movement, clothing, facial expressions, and environmental surroundings. Heuristics, including social presence, novelty, and authority, were all recognized by the final version of the VC, and creditably was established. The VC was named Agent Leveraging



XSL•FO
RenderX

Empathy for eXams (ALEX) and referred to as "brother-doctor," and participants stated "wanting to interact with ALEX over their regular doctor."

**Conclusions:** Involving Black men in the development of a digital health care intervention is critical. This population is burdened by cancer health disparities, and incorporating their perceptions in telehealth interventions will create awareness of the need to develop targeted messages for Black men.

***(JMIR Form Res 2021;5(12):e28709)*** doi:10.2196/28709

**KEYWORDS**

telehealth; digital health; eHealth; colorectal cancer; Black men; virtual human; technology; cancer screening; app; cancer; prevention; development

## Introduction

### Background

In the United States, colorectal cancer (CRC) is the second-leading cause of cancer-related death among both men and women. Despite improvements in incidence and mortality rates nationally, Black men bear a disproportionate burden of disease compared to the national average [1-3]. CRC incidence rates for non-Hispanic Black men in the United States are 58.3 per 100,000 compared to 46.9 per 100,000 among all men, with mortality rates of 25.9 per 100,000 compared to 17.7 per 100,000 among all men with CRC [2]. Multiple factors contribute to these inequities among Black men, including poorer access to preventive screening, aversion to colonoscopy, and limited knowledge of alternative screening modalities [4,5]. These inequities result in Black men reporting one of the lowest screening completion rates. To reduce CRC disparities among Black men, there remains a critical need to develop scalable CRC-screening promotion interventions that are tailored to meet the informational, cultural, and decision-making needs specific to Black men [6-10].

Home stool screening, through the fecal immunochemical test (FIT) kit, has been demonstrated as an effective alternative to colonoscopy [11-14]. FIT is a non-invasive stool test that requires no preparation, can be done in the comfort of the patient's home, and is inexpensive. Previous patient outreach studies have shown that when given a choice to screen [2] for CRC, screening rates are greater for FIT compared to colonoscopy among Black patients (43% vs 26.6%) [12].

### Digital Health Interventions and Virtual Clinicians

Digital health interventions, including telehealth, are provisions of health and prevention resources to patients irrespective of their geographic location. The long-term benefits of telehealth indicate reductions in avoidable health care service utilization and related costs [14-17]. Short-term benefits include the ability to disseminate and personalize health information to patients in real time [14-16]. Telehealth interventions can effectively promote CRC screening but have been underutilized in the promotion of FIT screening. This study aimed to combine telehealth with FIT to overcome historical barriers associated with accessing CRC-screening resources among Black men.

The evolution of telehealth interventions now encompasses highly customized interactions with virtual clinicians (VCs). The VC featured in this study is entitled "Meet ALEX." Agent Leveraging Empathy for eXams (ALEX) is an online virtual app using a visual representation of a health care provider that educates patients about the importance of CRC screening, identifies CRC-screening barriers, and provides a detailed description of the FIT kit. ALEX is interactive and replicates an in-person interaction using a combination of verbal and nonverbal behaviors. ALEX incorporates theory-driven conversational segments that include personalized risk information and barrier reduction strategies. Critically, ALEX can be tailored to reflect the demographic background (eg, race and gender) of the patient [18]. A growing body of literature has revealed demographic concordance between a health care provider and a racial/ethnic minority patient can increase adherence to recommended preventive testing and improve health outcomes [13,14]. Therefore, coupling a demographically concordant ALEX with FIT navigation offers promise to reduce CRC disparities among Black men. This method is only strengthened by incorporating Black men in the process of developing ALEX.

### MAIN Model Framework

A rigorous user-centered design (UCD) process was conducted to incorporate feedback from Black men to ensure ALEX was perceived as credible [19]. To provide a framework to interpret this data, we adopted the modality, agency, interactivity, and navigability (MAIN) model. The MAIN model offers an organizational framework to understand how a technology's affordances (ie, its capabilities) affect a user's perceptions of an interface [20]. The four broad categories of technological affordances are modality, agency, interactivity, and navigability. These affordances cue cognitive heuristics that lead to positive or negative credibility judgments after interacting with the technology [20-22].

In the MAIN model, each affordance is external to the content but affects the users' perception of the media's quality and credibility. The MAIN model has been applied to internet search engines, social media, video games, and virtual reality interfaces [23-25]. What is not always highlighted when the model is applied are the cues. Cues give way to some type of interaction between the designers and receivers in computer-mediated communications [26]. Cues are also known as effectors on users' interactions and perceptions [20,24,26-28]. In this study, we apply the MAIN model to ALEX, an online app that can be assessed by any device that can connect to the internet.



XSL•FO
RenderX

### Objectives

The purpose of this study was to (1) conduct a three-phase UCD process with Black men to develop and iteratively improve ALEX and (2) utilize the MAIN model to better understand heuristic cues identified by the Black male participants that lead to positive credibility judgments of ALEX before clinical dissemination.

## Methods

### Study Design

Black men participated across a three-phase UCD process, which included (1) needs investigation, (2) prototype development, and (3) evaluation [29,30]. A detailed overview of the UCD process has been published previously, which incorporated transdisciplinary expertise across communication science, computer science, a community advisory board, and a clinical team to develop the ALEX app [19].

### Participant Recruitment

Black men from the ages of 50 to 73 years were recruited between January 2017 and November 2018 using a purposive sampling strategy [20]. To be eligible for the study, participants had to be active patients in the university health care system; this inclusion criterion was established, given that the ALEX character would be used in a subsequent clinical trial involving a similar population to promote use of FIT. Eligible participants were identified via university-affiliated community engagement organizations and local senior centers from rural counties of North Florida. Participants were recruited by study staff through phone, email, and in-person methods [19].

Focus groups were conducted in person across various locations throughout the community. Each moderated focus group lasted between 60 and 90 minutes. To ensure rapport with the participants, the moderator was also a Black male. Institutional review board approval was granted for this study, and the participants provided informed consent prior to being audio-recorded. At the beginning of each interview, participants were provided information about the overall aim of the study and how the study team would incorporate participant feedback to help develop the Black male VC in the ALEX app. Each participant received a $40 gift card as compensation for their time and effort.

### Data Collection

Feedback was collected across 8 focus groups to iteratively and systematically develop the Black male VC. Instead of recording the script and developing the interactive VC animation for the first set of focus groups, we obtained initial feedback on the still images of the VC and then applied a script and engaged in character development. The UCD ensured the involvement and perceptions of the Black men throughout the entire process, systematically tailoring the VC based on their feedback. Each focus group discussion began with identifying the participants' familiarity, understanding, and knowledge of cancer, cancer screening, and virtual human technology. Participants in the first and second interviews viewed still images of the VC. Participants in the third, fourth, and fifth focus groups viewed iteration 1 of the VC prototype. Participants in the sixth and seventh focus groups viewed iteration 2 of the prototype, and participants in the eighth focus group provided feedback on the third and final iterations of the VC used in the ALEX app.

### Data Analysis

Audio recordings were transcribed verbatim and reviewed for accuracy. Transcripts were uploaded into NVivo (version 12.0; QSR International) to allow for electronic coding of the data. The data were analyzed using thematic analysis. Two coders (authors MV and DW) established an initial codebook with guidance from a senior author and expert in qualitative research (author JK). The codebook included a priori codes derived from the credibility literature and domains of inquiry from the interview guide. The full codebook is published in a previous publication [31].

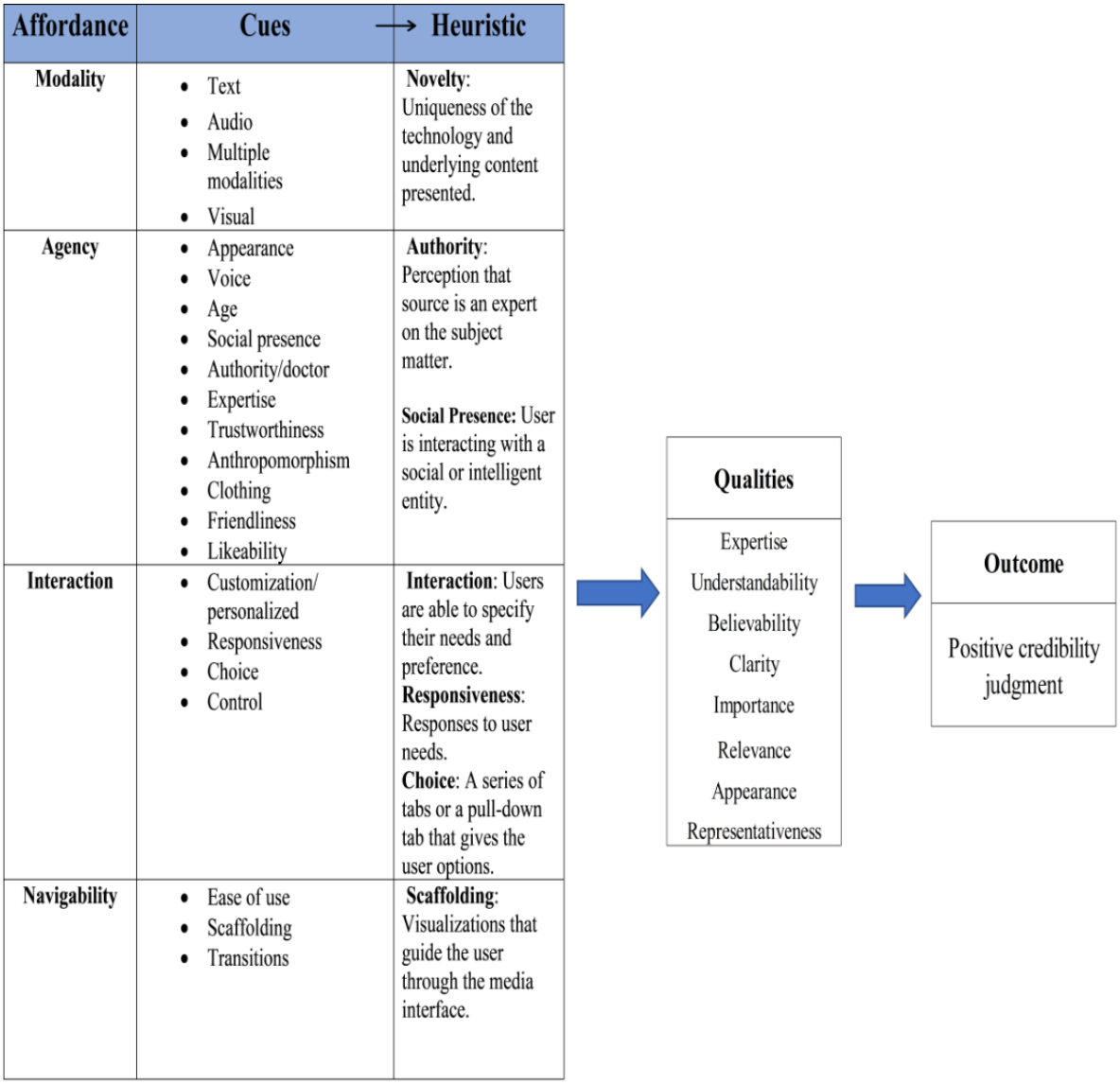
### Coding Procedures for the MAIN Model

Codes reflected theoretically informed components of credibility and guided a deductive coding approach for the presence of affordances, cues, and heuristics from the MAIN model, as seen in Figure 1. These cues prompted multiple heuristics such as social presence, authority, interaction, responsiveness, choice, control, and scaffolding, leading to an overall quality experience triggering enhanced credibility. The qualities associated with enhanced or positive credibility include uniqueness, trustworthiness, expertise, appearance, understandability, believability, clarity, importance, relevance, and representativeness. As shown in Figure 1, cues and heuristics were organized according to the affordance where modality examined how the content was conveyed through media (visual, text, audio, etc). Agency identified the source, that is, who or what was providing the content through the media interface (VC, text). Interactivity affordances identified cues that were associated with being in the presence of the digital media interface (the actions of clicking). Navigability examined the digital media interface's ability to guide the user through the content (dialog boxes, links, etc).



XSL•FO
RenderX

**Figure 1.** MAIN model affordance with the cues, heuristics, and qualities that lead to a positive credibility judgment for ALEX, the virtual health care assistant. MAIN: modality, agency, interactivity, and navigability; ALEX: Agent Leveraging Empathy for eXams.




During a line-by-line reading of the transcripts, emergent codes were added to the codebook through an inductive coding process. All codes were operationalized and assigned provisional definitions with exemplary quotes [31]. A formal coding framework was then established and used to determine interrater reliability (IRR). Two coders independently coded the transcript. An acceptable IRR was established by a κ coefficient above 0.8. The IRR was calculated using NVivo's coding comparison query. Initially, 16 coding discrepancies were documented and sent to a third reviewer (JK) for feedback. After 3 rounds of coding, low κ scores (below 0.5) were determined to be due to unitization issues rather than disagreement on coding [27]. To address this, transcripts were segmented by both coders by isolating text in the transcript using box frames to create a unit. Each unit was coded in its entirety to ensure the coders would capture the same information from beginning to end. In the fourth and final round of coding, κ was above 0.8 for all codes, and the rest of the transcripts were coded.

## Results

### Major Findings

Qualitative focus groups (included 7 focus groups and 1 individual interview) were conducted with a total of 25 Black men. The 1 individual interview occurred because 2 of the invited participants rescheduled on the day of the focus group. Therefore, he was the only participant during his focus group [19,29,32]. Demographic information of the men recruited can be found in Table 1. In summary, Black men were between the ages of 50 and 73 years (median age 61 [6.12]), with income ranging from less than $10,000 to over $100,000 annually. Their education levels ranged from eighth grade level to postsecondary and professional degrees.



XSL•FO
RenderX

**Table 1.** Demographic information about the focus groups of Black male participants (N=25).

| Demographics of focus groups of Black male participants (2017-2018) | n (%) |
|---|---|
| **Marital status** | |
| Married | 6 (24) |
| Divorced | 4 (16) |
| Separated | 6 (24) |
| Single | 4 (16) |
| Widowed | 1 (4) |
| Did not answer | 4 (16) |
| **Education level** | |
| Grades 1-8 | 1 (4) |
| Some high school | 2 (8) |
| High school graduate or General Educational Development (GED) certificate | 8 (32) |
| Technical, trade, or vocational school | 0 |
| Some college or associate degree | 5 (20) |
| College graduate (BS, BA, or other 4-year degree) | 3 (12) |
| Postgraduate training or professional school | 2 (8) |
| Did not answer | 4 (16) |
| **Employment status** | |
| Full-time | 2 (8) |
| Part-time | 5 (20) |
| Retired | 5 (20) |
| Volunteer | 1 (4) |
| Unable to work due to disability | 6 (24) |
| Unemployed | 3 (12) |
| Did not answer | 3 (12) |
| **Income level** | |
| Less than $10,000 | 7 (28) |
| $10,000-$19,000 | 5 (20) |
| $20,000-$34,999 | 6 (24) |
| $35,000-$49,999 | 1 (4) |
| $50,000-$74,999 | 0 |
| 75,000-$99,999 | 1 (4) |
| $100,000 or more | 1 (4) |
| Preferred not to answer | 2 (8) |
| Did not answer | 2 (8) |

Four of the focus groups were exclusively Black men, with three focus groups including a combination of Black and White men (Table 2). Feedback from the Black men was extracted and analyzed from the combination groups, as also shown in Table 2. For the developmental iterations of ALEX, the results are presented according to the affordances associated with the MAIN model: modality, agency, interaction, and navigability.



XSL•FO
RenderX

**Table 2.** Summary of focus groups and the number of Black men who participated.

| Black men only or combined Black and White men | Number of men (N=37), n (%) | Number of Black men (N=25), n (%) | Participant # of Black men | VC[a] version |
|---|---|---|---|---|
| Black men only (group 1) | 5 (14) | 5 (20) | 104-108 | Still images |
| Black men only (interview) | 1 (3) | 1 (4) | 109 | Still images |
| Black men only (group 3) | 4 (11) | 4 (16) | 14-17 | Version 1 |
| Black men only (group 4) | 2 (5) | 2 (8) | 18, 19 | Version 1 |
| Combined (group 5) | 4 (11) | 1 (4) | 73 | Version 1 |
| Combined (group 6) | 6 (16) | 4 (16) | 119-122 | Version 2 |
| Combined (group 7) | 7 (19) | 1 (4) | 126 | Version 2 |
| Combined (group 8) | 8 (22) | 7 (28) | 147-152, 154 | Version 3 |

[a]VC: virtual clinician.

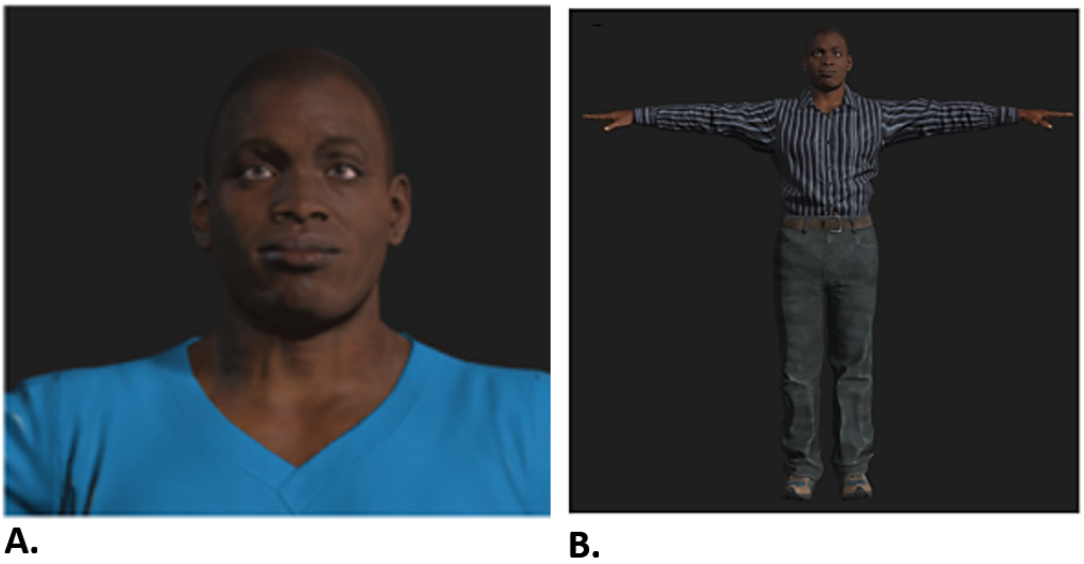
## Iteration 1: Still Images of ALEX

### *Modality*

Two focus groups viewed near and far printed images of the Black male VC, as shown in Figure 2. While participants reviewed the still images, moderators played voice clips and explained the plan to deliver the VC through multiple modalities (audio, visual, and text). Participants expressed strong support for the multimodal approach to intervention development. For example, one participant stated,

> *So, for people who can't read, that would be perfect.* [P107]

**Figure 2.** Near and far still images of the first iteration of the Black male ALEX character. A: the near iteration and B: the far iteration. ALEX: Agent Leveraging Empathy for eXams.




### *Agency*

Participants discussed agency in relation to how ALEX would function as a source of health information. Overall, participants expressed that obtaining health information from a VC was a new concept. These participants envisioned ALEX functioning like other trusted virtual apps, such as Google Maps or Siri, in which the VC would act like the Global Positioning System (GPS) for their CRC-screening journey. One participant noted,

> *I like that virtual human as long as they [are] just like maybe the GPS. They're gonna make sure you're right…And so that virtual human could be good if he's gonna make sure you stay on the path.* [P105]

For some, if they learned they were at high risk for CRC and needed to change modifiable behaviors to reduce their risk, as a result of information conveyed by ALEX, they would still seek a second opinion from their provider. One participant said,

> *I wouldn't accept it just because it was an app. You know, if the doctor told me that, if I knew my doctor and my nurse, and he told me that, you know, I'm feeling like he could probably sympathize, kind of urge me.* [P105]

### *Interaction*

Predicting the interaction with the app gave the men a sense of pride, knowing that it could be a resource that would help



XSL•FO
RenderX

facilitate a conversation with their doctors. One participant noted,

> *I could interact with it where I can be comfortable and go backward and forward and determining the severity of my situation, whereas to encourage me to go get medical assistance…do preventative maintenance.* [P106]

This cued the social presence–enhancing heuristic of the app, where the participants began to take ownership of it as a possible support system, as stated by one participant:

> *I'll try it out. I mean, if me and that app—I'd be at the hospital real fast and the thing about that app, for me, familiarized me with some of details how to explain to the doctor what I think is the symptoms that I'm coming in contact with.* [P104]

Participants discussed other cues associated with the interaction with ALEX, including perceptions of the customization or personalization of the VC. One participant said,

> *It asks me for a certain thing, and I give it to them. Then it shows me my risk factors. Now, if it shows me something that I would determine to be—I would take—I…I—no questions about it. I'd go to the doctor to make doggone sure it is—you know, either to curtail it or to prevent it. It doesn't matter.* [P106]

### *Navigability*

The means of accessing ALEX and navigating through the interface was important and played a role in increasing trust and credibility. Knowing that ALEX was designed, developed, and would be delivered from their health care system was important, as stated by one participant:

> *All my information is through email. Now, I would have to make sure— like you were saying, you got to make sure where it's coming from. Now, I'll see something that's associated with Shand's because that's where my doctors are at, and I click on it.* [P108]

## Iteration 2: First Video of ALEX

In iteration 2, 3 focus groups engaged with ALEX on a Samsung smartphone. In this iteration, participants viewed a computer software program that created an initial rendering of the VC, as shown in Figure 3. The VC could speak to participants, and the app contained subtitles corresponding to the speech.

**Figure 3.** Image of the first video of the virtual health care assistant, ALEX. ALEX: Agent Leveraging Empathy for eXams.

### *Modality*

Overall, the multimodalities were useful to participants. One participant said,

> *I liked that it was really, uh, it gave you all the information, it was on a personal level. So, man, woman, no matter who you are, you can listen to and watch the app. And that it has also, the, the words on the bottom. So, if you don't understand what the app is saying, you can also read it. That was really helpful.* [P16]

Credibility was essential to the success of the app so that patients would engage with it honestly. One participant stated,

> *We have to kind of win the confidence of the future patients that they [patients] will, though they are not in front of an actual doctor, they will tell the truth.* [P18]

### *Agency*

The virtual character had low visual fidelity with casual clothing and minimal breathing animations. The background of the character was a 3D model of a health care provider's office, as shown in Figure 3. Comments about agency reflected a range



XSL•FO
RenderX

of verbal and nonverbal cues, including the character's physical appearance. Specifically, ALEX's appearance was not "human-like" enough, with one participant commenting,

> *It looked like a mummy…that threw me off…it was futuristic.* [P73]

Others described the VC as "cartoonish" and even "scary," all of which diminished perceived credibility. In addition to comments about the physical appearance of the VC, other comments critiqued how the VC's clothing should look like a doctor's, with one participant commenting,

> *They should have [to] wear a white coat or something like that.* [P15]

The credibility of ALEX decreased as social presence was not established due to the disconnect of how unnatural the integration of his verbal and nonverbal gestures was perceived. One participant summarized perceptions of agency across multiple cues by saying,

> *Um, the eyes were a little narrow and just got distracted because they were so tiny. And then the language, the mouth, and his words were out of sync, too. So, if the mouth could be bigger and then in sync, it got distracting after a while, so I just decided to listen instead of looking at it. And maybe it's just from working in a hospital, but I wanted the doctor to be in a white coat. And you know, you guys probably went one way or another, but for me, I thought it just would have felt more natural.* [P17]

### *Interaction and Navigability*

There was a lot of discussion concerning the VC's appearance; yet, the Black male participants expressed "ease of use" when interacting and navigating the ALEX app, leading to interaction, responsiveness, and scaffolding heuristics. The focus groups that observed this iteration of the VC provided minimal feedback on interactivity but did discuss cues that reflected the responsive nature of the VC. One participant commented,

> *It was a lot of content, anticipated most of my questions…Yeah, I had a positive reaction to it.* [P17]

Participant perceptions surrounding the navigability of the intervention were coded in terms of ease of use, and in transitions where participants engaged with the VC through a mobile phone, participants found the app to be user friendly. According to one participant,

> *Um, I thought it was a good app. Interesting, a good app, um, seems to be pretty user friendly.* [P18]

However, another stated,

> *I thought it was pretty good information, it's pretty easy to use as you progress through it and use the tabs.* [P19]

## Iteration 3: Second Video of ALEX

Feedback from the previous focus groups that examined iteration 2 of ALEX led to a re-evaluation and development of the overall agent. This redevelopment of ALEX transpired over 9 months. The VC was further modified by a 3D artist to improve the appearance and match the health care providers' clothing. Nonverbal animations were recorded by recruiting gender-matched actors using the Vicon motion capture system. The virtual character's voice was recorded by professional voice talents who were race- and gender-matched with the virtual character. To make the environment like a real health care provider's office, we used a background image of a real clinical room. Changes were made to the Black male VC's eye shape and color, mouth, figure, hair, and posture, and more realistic movements were added to improve anthropomorphism. He was given props, such as a lab coat, name tag, and stethoscope, to convey an authoritative level of medical expertise, as shown in Figure 4. Facial expressions and hand gestures were added to the VC so that he could emotionally connect with the user.

**Figure 4.** The third iteration of ALEX. ALEX: Agent Leveraging Empathy for eXams.

XSL•FO
RenderX

### *Modality and Agency*

During the third iteration, although participants were largely positive about the inclusion of multiple modalities, concerns about a lack of coordination between modalities arose that could undermine the credibility of the app. In particular, participants were distracted by the fact that the audio and visual components (ie, lip synching) were mismatched. Participants made comments illustrating this idea, such as,

> *I stopped watchin' it and just started reading.* [*P* 124]
>
> *You can't listen and watch his mouth.* [P121]
>
> *Once I looked at it, that this thing was not human, it just made me turn off, although I listen at what it was sayin'. See, I couldn't see his body movements and the rhythm of his body…Is that thing computer generated? Because that's not a real person. I think it oughta have a real person do it.* [P119]

The negative feedback about the technology and the need to improve the quality of the interaction led to the moderator raising a question concerning adding a real doctor to the app to increase credibility and acceptability. This suggestion was well received by the participants:

> *I think that would be great.* [P119]

### *Interaction and Navigability*

The focus groups that observed the third irritation of ALEX provided minimal feedback on the interaction and navigability outside of not wanting to look at the VC. They expressed that navigation through the app was "simple" [P126].

## Iteration 4: Third Video of ALEX

The final focus group viewed the fourth iteration, which was the third and final version of the ALEX app, as shown in Figure 5, and the alterations were well received. Recommendations and comments from the previous focus groups concerning ALEX's movement and facial expressions were incorporated to increase the humanistic behavior of the VC. As the script was read by the hired voice actors, their facial expressions were recorded and the computer science team of the project applied the movements to ALEX, focusing on eye contact and mouth movements. The spatial usage of the patient room was also enhanced with a picture on the wall, as recommended by a participant, and an upgraded computational setup, and a patient bed was added for a more realistic experience, as shown in Figure 5. Near and far recordings were aligned with different parts of the script to assist with an enhanced interaction. In addition, a video of African American male physician introducing ALEX, virtual human technology, and the purpose of the intervention was added to the app to address previous concerns and comments addressing not using a person to deliver the intervention. This enabled changes in the script that allowed enabled ALEX to focus on CRC instead of having to explain the concept of a virtual human.

**Figure 5.** The final Iteration of ALEX. ALEX: Agent Leveraging Empathy for eXams.

XSL•FO
RenderX

### *Modality*

The feedback on the multiple modalities was positive, and the participants expressed satisfaction with being able to listen, watch, and read the material. One participant stated,

> *I like it because some people can't read fast enough and get it that way. They can watch a visual and soak it up better or watchin' the video. I thought it was great.* [P154]

Another participant stated that they preferred listening to the VC:

> *By me listening and listening and—I learn a lot about colon cancer, things that I didn't know.* [P148]

### *Agency*

By the fourth iteration, the VC had undergone substantial revision. Revision focused on improving the VC to invoke heuristics that included trustworthiness (through appearance and voice), expertise (clothing and visual field), and social presence (movement). Social presence, through natural movement, of the VC improved when verbal and nonverbal gestures were paired together. This was further strengthened when participants finally acknowledged gender and racial matching of the VC. The men's acceptance of ALEX reflected both outgroup (ie, a health care provider who was an authority) and in-group (social group membership through the same race and gender) characteristics, as expressed by one participant stating,

> *Now, me and the virtual guy, we partners…I wasn't interested until I saw virtual man.* [P152]

Other participants reflected this group membership combined with greater social presence by saying,

> *I thought he was pretty cool, since he was a brother…I mean, he gave some informative information, especially about the FIT. It was like watchin' a little cartoon like—Fat Albert-type situation, except he's talkin' more directly to a person or whatever.* [P150]

### *Interaction*

Credibility, likeability, identity, and social presence were all reached by the fourth iteration of the VC. Feedback concerning ALEX and interacting with ALEX was expressed as "natural," where the VC was life-like, as stated here:

> *I thought he was a real guy for a second.* [P154]
>
> *He did a lotta hand gestures, though, that's why it's so natural.* [P150]

The men described the interaction as straightforward and as a conversation where ALEX "don't beat around the bush neither" [P149] and "Tell it like it is" [P148]. The VC was seen as an expert and authoritative figure with a medical background:

> *I could talk to him better than I can a live doctor.* [P151]

### *Navigability*

There were no comments from participants regarding the navigability of the app.

## Content and Intended Behavior Outcomes from the ALEX App

One of the interesting outcomes from all focus groups was the attitude change toward screening after engaging with ALEX and learning more about CRC and the FIT kit. The content itself was novel to the men, and participants were interested in learning more, talking to their doctors, and sharing the ALEX app with others. One participant who viewed iteration 3 commented,

> *I'm gonna recommend that website. I'm gonna show you what I learned.* [P109]

It was important for the men to get the message about CRC screening out to those who could benefit:

> *Everybody needs that message. We're talkin' about it today cause it's important. You know what I mean? I don't want nobody to go without hearin' the message. I would wanna share it…It's stuff you might not get from your doctor.* [P126]

Specifically, the VC app was seen as supplemental information that may not be obtained otherwise.

The novelty of the app was engaging and helped men see the benefits of CRC screening:

> *Like I said, about the FIT, I didn't know, okay? I'm gonna get it done. I wanna get it done…In a heartbeat.* [P154]

The demonstration of how to use the FIT kit was well received and informed one of the participants who had the kit at home but did not know how to use it"

> *In the application, when you put the thing across the seat—cause I didn't know that. That's why I haven't really been botherin' with the thing. But now, I might try and it put across the seat…If I find out, from the stool, that I need to go and get the camera up in my bottom, then that's okay—I needed to know cause I got the test at home, but I just didn't know how to do it.* [P152]

For other men, this information was viewed as novel:

> *I didn't know about the FIT. Most men are embarrassed about goin', gettin' colonoscopies done.* [P154]

In addition, it provided a new way of looking at CRC screening:

> *This video, this app will open people eyes. It opened mine.* [P148]

# Discussion

## Principal Findings

In this study, Black men informed the development of a Black male VC, ALEX, that provided education on CRC, at-home screening recommendations, and risk factors. Participants



XSL•FO
RenderX

influenced the visual, audio, and textual content of this app, ensuring it would be attractive and informative for future men who would interact with this app as part of a clinical trial.

The MAIN model was applied to this study to assess the credibility of ALEX. Feedback from the participants was categorized according to the four affordances.

One unique aspect of this study was the overlapping heuristic cues identified. Traditionally specific heuristic cues were associated with specific affordances [20,21,27,28]. Some heuristics such as social presence and identity were found to be associated or cued by both agency and interaction affordances in different mediums [20,21,27,28]. In this study, multiple cues triggered overlapping heuristics that led to overall enhanced user perception and credibility judgement with the VC and usage of the FIT kit. The converging cues triggered multiple heuristics associated with multiple affordances. The heuristics then led to positive qualities, including expertise, understandability, believability, clarity, importance, relevance, appearance, and representative, which were all referred to by the focus group participants at one time or another throughout this study. For instance, the multiple modalities associated with the modality affordance triggered the being there, realism, and novelty heuristics and also triggered social presence and identity when the men called ALEX their “brother-man” or would rather interact with ALEX instead of their doctor. They perceived him as a member of their inner circle because multiple modalities were provided. They could visually see ALEX and watch his movements, eye contact, and mannerisms. They could also hear the audio and his words, diction, and tone. This increased the credibility of the ALEX app.

Through their interaction with the app, the anthropomorphism of ALEX improved the most throughout the focus groups. The realism and authority heuristics were cued from the final corrections needed to establish anthropomorphism. ALEX’s movement was one of the largest distractions for the focus group members. Once the movement of the VC perfectly aligned with the audio, the men were able to focus more on the content of the app, as shown in the results.

Since the goal of this app is to encourage the user to make a positive decision regarding their health and screening for CRC, a novel affordance emerged, the content itself. The content examined the characteristics associated with conveying the right message. The men spoke of accuracy of the content related to having gone through FIT previously. The importance and understandability of the content motivated the men, and they wanted to share the app or obtain more information from their doctors. All possible cues (understandability, clarity, accuracy) for the content are examples of converging cues, where they lead to authority, helper, own-ness, and novelty heuristics associated with other affordances.

ALEX was seen both as an authority or a helper by the men, where it could serve as a resource at the doctor’s office. Others saw it as an intelligent entity and would rather engage with the app instead of a doctor. The own-ness heuristic, defined as the ability to adapt the content to the perceived interest of the users, enabled the men to make recommendations for the content and identify things that were confusing to them. We were then able to address those concerns and personalize and tailor the information for the intended users as part of the clinical trial.

The FIT kit as an option for CRC screening was highly novel to most of the focus group participants. Multiple studies have explored the methods of promoting screening via at-home fecal testing within underserved populations [6,7,33-36]. A systematic review published in 2018 identified 27 unique CRC-screening interventions using FIT or the fecal occult blood test with low-income and rural populations in different settings [36]. The authors identified the most effective studies focused on increasing community access to fecal testing, increasing or providing delivery of testing kits, developing tailored decision aids promoting screening, or interacting with patient navigators, clinicians, or community members to increase understanding of fecal testing [36]. The ALEX app captures all of those efforts in one setting.

ALEX serves as a virtual patient navigator that was designed by the targeted population, promotes screening, and provides delivery of the testing kit. Overall, the novelty and coolness of the ALEX app promoting FIT kits was well received and led to increased interest in screening, which is the goal of the app. The Black men in our study were not only interested in the FIT kit after interacting with ALEX but also wanted to share the app. This interface was liked and found trustworthy by the Black male focus group participants.

## Conclusion

In this study, the feedback from Black men helped enhance the design of the ALEX app. Future implications of the ALEX app include involvement as part of a clinical trial to deliver tailored messages promoting FIT uptake. This study entailed a great deal of reflexivity on the part of the research team. The current social climate has made direct conversations around race challenging, which is one reason we think this work is so important. The research team is racially and ethnically diverse, and we engaged in weekly conversations about both the conduct of the research study as well as the larger context. We also drew on the experience and wisdom of a racially and ethnically diverse community advisory board. We have built this relationship over the past 4 years, which enabled us to have direct and meaningful conversations about the methods as well as the results.

As digital health care interventions continue to advance, involving specific populations in the design and development of these technologies is critical. Targeted interventions, such as the one in this study, can lay the groundwork for engaging Black men in science. Long term, it can increase participation in health care interventions by promoting trust, inclusion, and responsiveness.



XSL•FO
RenderX

## Conflicts of Interest

None declared.

## References


1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. American Cancer Society. Colorectal Cancer Facts & Figures 2017–2019. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2017-2019.pdf [accessed 2019-01-22]
3. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer 2019 Apr 15;144(8):1941-1953 [FREE Full text] [doi: 10.1002/ijc.31937] [Medline: 30350310]
4. May FP, Bromley EG, Reid MW, Baek M, Yoon J, Cohen E, et al. Low uptake of colorectal cancer screening among African Americans in an integrated Veterans Affairs health care network. Gastrointest Endosc 2014 Aug;80(2):291-298 [FREE Full text] [doi: 10.1016/j.gie.2014.01.045] [Medline: 24674351]
5. Myint A, Aby ES, May F. Getting colorectal cancer screening right: a teachable moment. JAMA Intern Med 2019 Oct 01;179(10):1417-1418 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3457] [Medline: 31424479]
6. Xirasagar S, Li Y, Burch JB, Daguisé VG, Hurley TG, Hébert JR. Reducing colorectal cancer incidence and disparities: performance and outcomes of a screening colonoscopy program in South Carolina. Adv Public Health 2014;2014:1-13 [FREE Full text] [doi: 10.1155/2014/787282] [Medline: 25705719]
7. Naylor KB, Fritz CD, Kim KE. Comparing screening colonoscopy quality measures in navigated uninsured and insured patient populations. Gastroenterology 2014 May;146(5):S-70. [doi: 10.1016/s0016-5085(14)60254-7]
8. Hunleth JM, Steinmetz EK, McQueen A, James AS. Beyond adherence: health care disparities and the struggle to get screened for colon cancer. Qual Health Res 2016 Jan;26(1):17-31 [FREE Full text] [doi: 10.1177/1049732315593549] [Medline: 26160775]
9. Rogers CR, Goodson P. Male role norms, knowledge, attitudes, and perceptions of colorectal cancer screening among young adult African American men. Front Public Health 2014;2:252 [FREE Full text] [doi: 10.3389/fpubh.2014.00252] [Medline: 25506049]
10. James AS, Campbell MK, Hudson MA. Perceived barriers and benefits to colon cancer screening among African Americans in North Carolina: how does perception relate to screening behavior? Cancer Epidemiol Biomarkers Prev 2002 Jun;11(6):529-534 [FREE Full text] [Medline: 12050093]
11. Sly JR, Edwards T, Shelton RC, Jandorf L. Identifying barriers to colonoscopy screening for nonadherent African American participants in a patient navigation intervention. Health Educ Behav 2013 Aug;40(4):449-457 [FREE Full text] [doi: 10.1177/1090198112459514] [Medline: 23086556]
12. Taber JM, Aspinwall LG, Heichman KA, Kinney AY. Preferences for blood-based colon cancer screening differ by race/ethnicity. Am J Health Behav 2014 May;38(3):351-361. [doi: 10.5993/AJHB.38.3.4] [Medline: 24636031]
13. Thorpe RJ, Bowie JV, Wilson-Frederick SM, Coa KI, Laveist TA. Association between race, place, and preventive health screenings among men: findings from the exploring health disparities in integrated communities study. Am J Mens Health 2013 May;7(3):220-227 [FREE Full text] [doi: 10.1177/1557988312466910] [Medline: 23184335]
14. Chun FK, Suardi N, Perrotte P, Lebeau T, Guay J, Benayoun S, et al. Screening men for prostate cancer and colorectal cancer: is practice evidence-based? Can J Urol 2007 Dec;14(6):3727-3733. [Medline: 18163923]
15. Escriva Boulley G, Leroy T, Bernetière C, Paquienseguy F, Desfriches-Doria O, Préau M. Digital health interventions to help living with cancer: a systematic review of participants' engagement and psychosocial effects. Psychooncology 2018 Dec 24;27(12):2677-2686. [doi: 10.1002/pon.4867] [Medline: 30152074]
16. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
17. Wirth MD, Brandt HM, Dolinger H, Hardin JW, Sharpe PA, Eberth JM. Examining connections between screening for breast, cervical and prostate cancer and colorectal cancer screening. Colorectal Cancer 2014 Jun;3(3):253-263 [FREE Full text] [doi: 10.2217/crc.14.18] [Medline: 25143785]
18. Sanders MS, Ramasubramanian S. An examination of African Americans' stereotyped perceptions of fictional media characters. Howard J Commun 2012 Jan;23(1):17-39. [doi: 10.1080/10646175.2012.641869]
19. Griffin L, Lee D, Jaisle A, Carek P, George T, Laber E, et al. Creating an mHealth app for colorectal cancer screening: user-centered design approach. JMIR Hum Factors 2019 May 08;6(2):e12700 [FREE Full text] [doi: 10.2196/12700] [Medline: 31066688]




XSL•FO
RenderX

20. Sundar S. Digital media, youth, and credibility. In: Metzger MJ, Flanagin AJ, editors. The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. Cambridge, MA: The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning; 2008:73-100.
21. Sundar SS, editor. The Handbook of the Psychology of Communication Technology. Chichester, West Sussex, UK: Wiley Blackwell; 2015.
22. Balakrishnan B, Sundar SS. Where am i? How can I get there? Impact of navigability and narrative transportation on spatial presence. Hum Comput Interact 2011 Sep 02:161-204. [doi: 10.1080/07370024.2011.601689]
23. Oh J, Sundar SS. What Happens When You Click and Drag: Unpacking the Relationship between On-Screen Interaction and User Engagement with an Anti-Smoking Website. Health Commun 2020 Mar;35(3):269-280. [doi: 10.1080/10410236.2018.1560578] [Medline: 30618306]
24. Li R, Sundar SS. Can Interactive Media Attenuate Psychological Reactance to Health Messages? A Study of the Role Played by User Commenting and Audience Metrics in Persuasion. Health Commun 2021 Feb 23:1-13. [doi: 10.1080/10410236.2021.1888450] [Medline: 33622093]
25. Waddell F, Sundar S, Auriemma J. Can customizing an avatar motivate exercise intentions and health behaviors among those with low health ideals? Cyberpsychol Behav Soc Netw 2015 Nov;18(11):687-690. [doi: 10.1089/cyber.2014.0356] [Medline: 26406804]
26. Xu K, Liao T. Explicating cues: a typology for understanding emerging media technologies. J Comp-Mediat Commun 2019 Dec 17;25(1):32-43. [doi: 10.1093/jcmc/zmz023]
27. Sundar SS, Kim J. Interactivity and persuasion. J Interact Advert 2005 Mar;5(2):5-18. [doi: 10.1080/15252019.2005.10722097]
28. Sundar S, Jia H, Waddell F, Huang Y. Toward a theory of interactive media effects (TIME) four models for explaining how interface features affect user psychology. In: The Handbook of the Psychology of Communication Technology. Chichester, West Sussex, UK: Wiley-Blackwell; 2015.
29. Sewitch MJ, Fallone CA, Ghali P, Lee GE. What patients want in a smartphone app that supports colonoscopy preparation: qualitative study to inform a user-centered smartphone app. JMIR Mhealth Uhealth 2019 Jul 02;7(7):e12242 [FREE Full text] [doi: 10.2196/12242] [Medline: 31125310]
30. Wachtler C, Coe A, Davidson S, Fletcher S, Mendoza A, Sterling L, et al. Development of a mobile clinical prediction tool to estimate future depression severity and guide treatment in primary care: user-centered design. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e95 [FREE Full text] [doi: 10.2196/mhealth.9502] [Medline: 29685864]
31. Vilaro MJ, Wilson-Howard DS, Griffin LN, Tavassoli F, Zalake MS, Lok BC, et al. Tailoring virtual human-delivered interventions: a digital intervention promoting colorectal cancer screening for Black women. Psychooncology 2020 Dec;29(12):2048-2056 [FREE Full text] [doi: 10.1002/pon.5538] [Medline: 32893399]
32. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013 Jul 02;1(2):e8-e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
33. Babikian R, Chang M. Colonoscopy education among African-American males in the primary care setting of Bronx County, New York. Gastroenterol Nurs 2015;38(5):337-341. [doi: 10.1097/SGA.0000000000000129] [Medline: 26422268]
34. Collazo TH, Jandorf L, Thelemaque L, Lee K, Itzkowitz SH. Screening colonoscopy among uninsured and underinsured urban minorities. Gut Liver 2015 Jul 23;9(4):502-508 [FREE Full text] [doi: 10.5009/gnl14039] [Medline: 25287165]
35. Tsai M, Xirasagar S, de Groen PC. Persisting racial disparities in colonoscopy screening of persons with a family history of colorectal cancer. J Racial Ethn Health Disparities 2018 Aug;5(4):737-746. [doi: 10.1007/s40615-017-0418-1] [Medline: 28812255]
36. Dimou A, Syrigos KN, Saif MW. Disparities in colorectal cancer in African-Americans vs Whites: before and after diagnosis. World J Gastroenterol 2009 Aug 14;15(30):3734-3743 [FREE Full text] [doi: 10.3748/wjg.15.3734] [Medline: 19673013]

## Abbreviations

**ALEX:** Agent Leveraging Empathy for eXams
**CRC:** colorectal cancer
**FIT:** fecal immunochemical test
**GPS:** Global Positioning System
**IRR:** interrater reliability
**MAIN:** modality, agency, interactivity, and navigability
**UCD:** user-centered design
**VC:** virtual clinician



XSL•FO
RenderX

*Edited by G Eysenbach; submitted 16.03.21; peer-reviewed by C Richardson; comments to author 07.04.21; revised version received 21.04.21; accepted 06.10.21; published 24.12.21.*

*Please cite as:*
*Wilson-Howard D, Vilaro MJ, Neil JM, Cooks EJ, Griffin LN, Ashley TT, Tavassoli F, Zalake MS, Lok BC, Odedina FT, Modave F, Carek PJ, George TJ, Krieger JL*
*Development of a Credible Virtual Clinician Promoting Colorectal Cancer Screening via Telehealth Apps for and by Black Men: Qualitative Study*
*JMIR Form Res 2021;5(12):e28709*
*URL: https://formative.jmir.org/2021/12/e28709*
*doi:10.2196/28709*
*PMID:34780346*

©Danyell Wilson-Howard, Melissa J Vilaro, Jordan M Neil, Eric J Cooks, Lauren N Griffin, Taylor T Ashley, Fatemeh Tavassoli, Mohan S Zalake, Benjamin C Lok, Folakemi T Odedina, Francois Modave, Peter J Carek, Thomas J George, Janice L Krieger. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Acceptability of an mHealth App for Youth With Substance Use and Mental Health Needs: Iterative, Mixed Methods Design

Zachary Adams[1], PhD; Miyah Grant[2], PsyD; Samantha Hupp[1], MS; Taylor Scott[3], BS; Amanda Feagans[1], MS; Meredith Lois Phillips[1], MS; Kristina Bixler[1], PhD; Phani Teja Nallam[3], MS; Dorothy La Putt[3], BFA

[1]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, United States

[2]Department of Psychology, University of Indianapolis, Indianapolis, IN, United States

[3]Department of Psychiatry, Indiana University-Purdue University Indianapolis, Indianapolis, IN, United States

**Corresponding Author:**
Zachary Adams, PhD
Department of Psychiatry
Indiana University School of Medicine
340 W 10th St.
Indianapolis, IN, 46202
United States
Phone: 1 317 278 0591
Email: zwadams@iu.edu

## Abstract

**Background:** Treating substance use disorders (SUDs) during adolescence can prevent adult addiction and improve youth outcomes. However, it can be challenging to keep adolescents with SUDs engaged in ongoing services, thus limiting potential benefits. Developmentally appropriate tools are needed to improve treatment engagement during and between sessions for youth with SUDs and mental health disorders. Mobile health apps may augment or replace psychotherapy components; however, few have been developed specifically for youth with SUDs following user-guided design principles, which may limit their appropriateness and utility. Formative research on acceptability to intended end users is needed before the efficacy of such tools can be examined.

**Objective:** This study involves user-centered, iterative development and initial user testing of a web-based app for adolescents with SUDs and mental health concerns.

**Methods:** Adolescents aged 14 to 17 years with past-year involvement in outpatient psychotherapy and behavioral health clinicians with adolescent SUD treatment caseloads were recruited. Across 2 assessment phases, 40 participants (alpha: 10 youths and 10 clinicians; beta: 10 youths and 10 clinicians) viewed an app demonstration and completed semistructured interviews and questionnaires about app content and functionality.

**Results:** Participants expressed positive impressions of the app and its potential utility in augmenting outpatient therapy for youth with SUDs and mental health concerns. Noted strengths included valuable educational content, useful embedded resources, and a variety of activities. Adolescents and clinicians favored the app over conventional (paper-and-pencil) modalities, citing convenience and familiarity. The app was found to be user-friendly and likely to improve treatment engagement. Adolescents suggested the inclusion of privacy settings, and clinicians recommended more detailed instructions and simplified language.

**Conclusions:** The novel app developed here appears to be a promising, acceptable, and highly scalable resource to support adolescents with SUDs and mental health concerns. Future studies should test the efficacy of such apps in enhancing adolescent behavioral health treatment engagement and outcomes.

***(JMIR Form Res 2021;5(12):e30268)*** doi:10.2196/30268

**KEYWORDS**

mobile health; user-centered design; adolescents; substance use disorders; mental health; mHealth; cognitive behavioral therapy; homework; technology acceptance model; trauma; mobile phone



XSL•FO
RenderX

## Introduction

### Background

Co-occurring substance use and mental health disorders are common in adolescence. Approximately 916,000 (3.7%) adolescents aged 12 to 17 years in the United States meet the criteria for ≥1 substance use disorder (SUD) [1]. Approximately 38% of 12th graders experimented with illicit drug use within the past year, whereas 47% reported using illicit drugs at least once during their lifetime [2]. Compared with adults, adolescents with SUDs are more likely to have a co-occurring mental health disorder [3], with 60% to 75% of adolescents exhibiting SUD meeting criteria for ≥1 disruptive behavior disorders, mood disorders, anxiety disorders, traumatic stress disorders, or other psychiatric disorders [1,4,5].

Despite the high rates of behavioral health disorders among youth, few receive appropriate health treatment services. In 2018, only 1 in 10 youths with SUDs received behavioral health treatment [6]. Individuals with co-occurring disorders typically experience greater symptom severity, require greater service use throughout treatment, and demonstrate poorer treatment outcomes compared with youth with single disorders [7,8]. Although treatment interventions for adolescents with SUDs show promising long-term outcomes, approximately 40% of adolescents discontinue prematurely before completing their treatments [9]. Collectively, these trends point to a need for better strategies for improving youth engagement in treatment services.

An additional barrier to care is that few intervention resources have been developed that specifically address improving care for youth with comorbid disorders [10]. Moreover, clinical service systems are often disjointed and address substance use and mental health separately [11]. Treatments meant to address co-occurring disorders often incorporate well-established, evidence-based principles often rooted in a cognitive behavioral therapy (CBT) framework. Such treatments emphasize the importance of identifying specific goals, understanding the antecedents and consequences of substance use, teaching new coping skills (eg, substance refusal, emotion regulation and relaxation, communication skills, and cognitive restructuring), progress monitoring, and practicing strategies in and between sessions [12-14]. Increasing patient involvement in homework exercises may be especially valuable, as more consistent homework completion in CBT is associated with greater skill mastery and more improvements in key clinical domains [14-16].

These components can be challenging to implement in conventional treatment contexts. For instance, youth often struggle to complete homework tasks as recommended in CBT [15,17]. Moreover, many clinicians lack confidence in treating co-occurring disorders, which can reduce fidelity to best practice models [18]. Thus, there is a need for scalable, sustainable resources that are not only engaging for youth but also helpful for providers in delivering evidence-based interventions to this population in line with best practice guidelines and principles.

Mobile health (mHealth) solutions that leverage remote functionality of smartphones and other devices are an increasingly feasible strategy for improving behavioral health treatments for adolescents [19], especially given the very high smartphone ownership and accessibility among youth, with 95% of teenagers reporting access to smartphone devices in 2019 [20]. There is also evidence that adolescents are open to using smartphone-delivered tools to support recovery and relapse prevention as part of SUD treatment [21]. Such tools, whether delivered via mobile apps or SMS text messaging, have shown promise in improving treatment adherence and engagement while also expanding access to care [22-24]. mHealth solutions can directly address the key treatment components of CBT. For instance, regarding homework, apps may send automated calendar reminders, present individualized guidance and educational content, and offer real-time feedback to users about their responses or behavior in ways that complement clinician-delivered content [19,25]. Although a growing number of mobile apps have emerged for mental and physical health, few are specifically designed to address the needs and preferences of adolescents with SUDs and co-occurring disorders [23,26].

When developing mHealth apps, the current best practice is to follow a user-centered, iterative design process to optimize usability and effectiveness for intended end users [27-29]. In this framework, both users and experts test the app and provide feedback on its relevance and functionality, which allows the tool to be tailored to the user's needs while still adhering to theoretically or empirically supported principles [28]. This approach can provide valuable insight into which features are most important to users, such as personalization, autonomy, simplicity, and informativeness [30]. This process often involves a mixed methods approach to assessment, whereby users interact with the app or view a demonstration and are then given opportunities to share their reactions about the design, features, and functions of the app via standardized survey instruments and interviews.

As the goal of mHealth interventions is to allow users to change their behavior, both in terms of SUD and mental health–related behaviors but also with respect to using the app and adhering to treatment recommendations, the app development and evaluation process may also be guided by prominent behavior change theories. One such framework is an extension of the theory of reasoned action called the technology acceptance model (TAM), which posits that acceptance and attitude toward technology are directly influenced by perceived usefulness (PU) and perceived ease of use (PEOU) in achieving the intended goal (eg, delivering key treatment components such as homework activities or symptom monitoring) [31]. Such formative work to determine the acceptability and appropriateness of an app is critical before later tests of efficacy and effectiveness in improving clinical outcomes.

### Objective

The purpose of this study is to develop and perform initial usability pilot-testing of a mobile app designed to augment outpatient behavioral health treatment for adolescents with SUDs and co-occurring mental health concerns. Interactive



XSL•FO
RenderX

content was rooted in evidence-based treatment principles, designed to appeal to youth, and evaluated following an iterative, user-centered design strategy in which both adolescents with SUDs and clinicians who treat youth with SUDs were involved to maximize relevance to treatment stakeholders.

## Methods

### Recruitment

Participants were recruited for two phases of this formative acceptability study: alpha testing (10 adolescents and 10 clinicians) and beta testing (10 youths and 10 clinicians). This sample size was initially selected to help ensure the collection of diverse reactions to the app from the intended end user groups at different points in the app development process. Participants viewed an app demonstration and completed semistructured interviews and questionnaires about app content and functionality.

Adolescent participants aged 14 to 17 years (alpha: mean 15.3, SD 1.0 years, 6/10, 60% girls and 3/10, 30% boys; beta: mean 15.9, SD 1.0 years, 8/10, 80% girls and 2/10, 20% boys) had current or past-year involvement in outpatient psychotherapy for ≥1 SUDs per self- and caregiver-report at the time of eligibility phone screening. Car, relax, alone, forget, friends, and trouble substance use screeners were also administered during phone screening to verify likely SUD status (all participants scored ≥1, consistent with published cutoff recommendations) [32-34]. Comprehensive diagnostic evaluations were outside the scope of this study and were not performed.

Clinician participants were eligible if they had an active adolescent SUD treatment caseload (range: alpha, 20%-100% of caseload; beta, 10%-100% of caseload) and had worked with ≥3 adolescents with co-occurring SUDs and mental health conditions in the past 5 years. Clinician participants (all female; age—alpha: mean 43.3, SD 7.5 years; beta: mean 37.8, SD 12.1 years) reported varied credentials (licensed clinical social worker: 6/20, 30%; licensed social worker: 4/20, 20%; licensed medical health counselor: 3/20, 15%; PhD: 2/20, 10%; PsyD: 2/20, 10%; other: 3/20, 15%) and years of experience (mean 11.6, SD 7.5 years). Both adolescents and clinicians were recruited from community and mental health and SUD treatment facilities, academic health system clinical research registries, community advertisements, and word of mouth. The procedures were approved by the university institutional review board. All participants completed consent and assent procedures during the enrollment process.

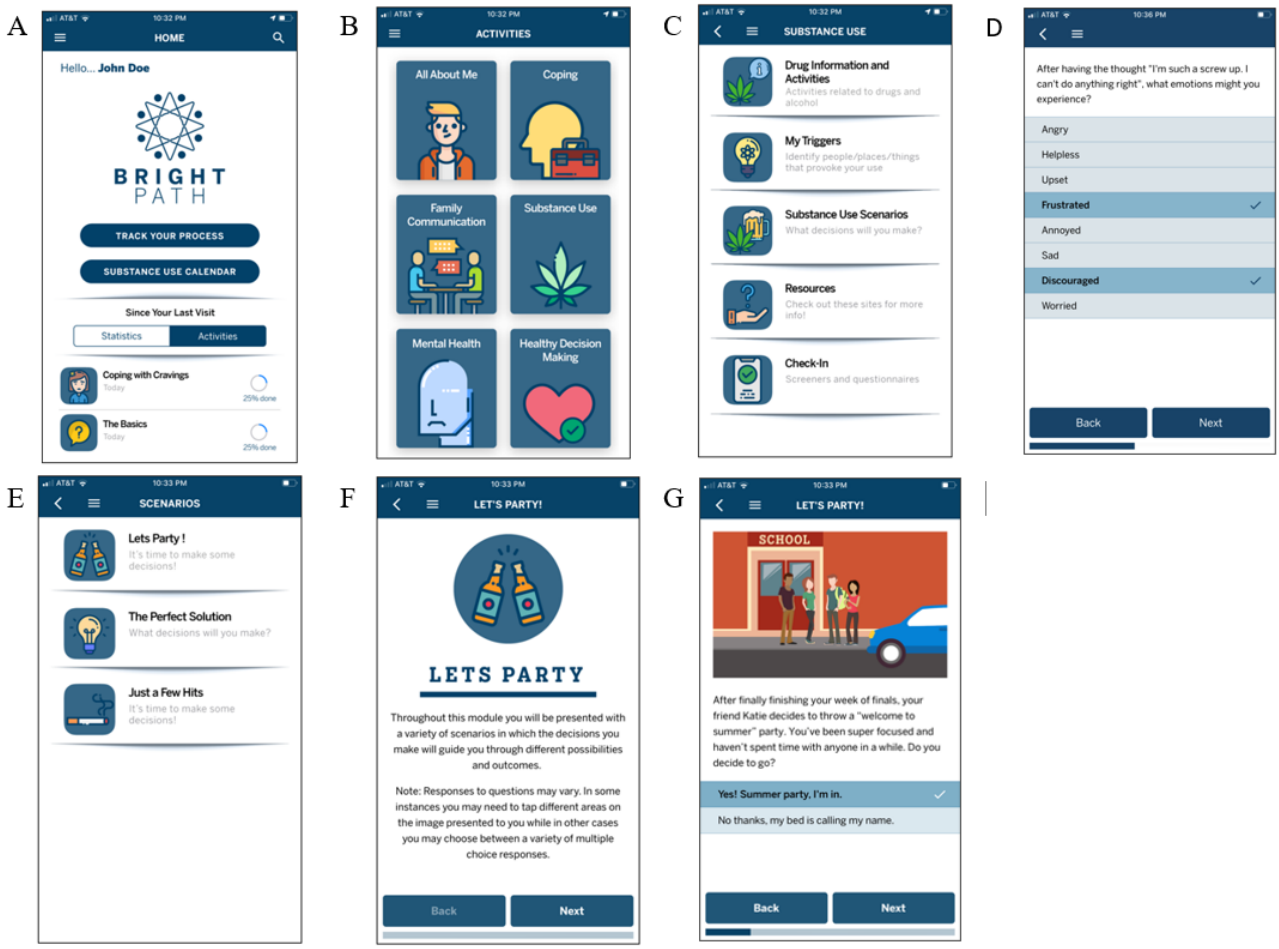
### mHealth App: Bright Path

The novel, web-based mHealth app developed here was designed to present educational content and interactive games and activities that address common factors associated with substance use and mental health disorders in youth. The content is arranged into several sections (coping skills, substance use, mental health, family communication, and healthy decision-making), which are accessible from a menu screen. Within each section, users are presented with a menu of activities to choose from. For instance, within the *substance use* content area, users can access activities where they can perform tasks such as identifying their personal cues and triggers for substance use, completing substance use assessments (screening tools and self-report of recent substance use), practicing safe decisions in high-risk hypothetical scenarios, or learning about the effects of different types of substances. Within the *coping* section, activities address cognitive behavioral skills such as challenging automatic thoughts, selecting and scheduling prosocial activities, and learning about connections between thoughts, feelings, and behaviors. Throughout the app, each activity begins with basic instructions to help users know how to interact with the content. Several activities also include feedback slides that provide additional information about correct and incorrect answers or provide encouragement for continued progress. The app also includes an *all about me* section where users can enter basic information about themselves and customize certain features such as their name and picture. Activities are presented in various formats. Some are presented as surveys, some as web-based *card decks* (ie, question on one side and answer on the other side), and some as web-based *choose your own adventure* style scenarios. Figure 1 presents screenshots from the app depicting the general look and feel as well as the organization of the app.



XSL•FO
RenderX

**Figure 1.** Screenshots from the Bright Path app. (A) Home screen with progress monitoring dashboard; (B) main menu of content topics; (C) first page of substance use–related activities menu; (D) sample of multiselect items included in an interactive coping exercise; (E) menu of scenario activities where users are instructed to read vignettes and make choices that advance the story; each choice either increases or decreases risk for substance use or negative health outcomes; (F) sample instruction screen from a scenario practice activity; (G) sample screen from a scenario practice activity with illustration and text accompanied by response choices. Feedback is presented to users based on their selections, and the story advances.




## Measures

### *Qualitative Interview*

In both the alpha and beta phases of the study, the mobile app *e-tools* were demonstrated in person on smartphones or through videoconferencing software. A semistructured interview was conducted using the think-aloud technique, a technique that has been routinely and successfully implemented in the usability evaluations of other web-based interventions [35,36]. Participants' responses as they were guided through the app were audio recorded for transcription and analysis following the visit. Participants reviewed 1 of 2 subsets of approximately 8 interactive components within the app. Each subset component was selected on an alternating basis whereby half the participants viewed the first subset, and the other half viewed the second subset. An average of 5 to 10 minutes was dedicated to reviewing each component.

### *Quantitative Rating Instruments*

Assessment and refinement of the mobile app were guided by the TAM, an extension of the theory of reasoned action. The two key factors emphasized by the TAM are (1) PU, which is the degree to which tools will accomplish the goal of enhancing treatment, and (2) PEOU, which encompasses the ease of navigation, technical problems, and reactions to the overall *look and feel* of the interface. PU and PEOU are theorized to predict actual system use through the measurement of attitudes toward use and behavioral intentions to use. Additional items assessed perceived ease of learning and overall satisfaction. A 46-item questionnaire administered to both youth and clinicians also assessed respondents' perceived self-efficacy when using mobile devices as well as concerns related to the use of technology in treatment (eg, privacy and social acceptability). Subscales measuring PU (13 items; sample item: Using this app in treatment would help my patients accomplish tasks and goals more quickly), PEOU (14 items; sample item: I would find it easy to get this app to do what I want it to do), perceived ease of learning (4 items; sample item: It would be easy for me to become skillful using this app), and satisfaction (14 items; sample item: Using the app is a good idea) were calculated. These measures demonstrated internal consistency (PU, providers α=.89, youth α=.83; PEOU, providers α=.88, youth α=.92; perceived ease of learning, providers α=.87, youth α=.81; and satisfaction, 14 items, providers α=.86, youth α=.87).



XSL•FO
RenderX

### *Name and Logo Design*

During the alpha phase of the study, youth and providers were presented with an array of 16 potential logos and app names generated through a series of brainstorming sessions within the clinical research team and in partnership with experts in graphic design. Participants were directed to select and rank their *top 5* preferred logo options. The most highly rated name (Bright Path) and logo across the youth and clinician participants were then selected and used throughout the beta phase of the study. The logo is depicted in Figure 1.

### Statistical Analysis

Qualitative data obtained during the interviews were transcribed and summarized. A total of 2 independent reviewers analyzed and interpreted the transcriptions, identifying and coding for common themes across participants and extracting illustrative quotes for each theme. Participant feedback that was collected during alpha testing guided refinements for beta testing to improve PU and PEOU. Mann–Whitney *U* tests were performed to test the hypothesis that PU and PEOU scores at beta testing were higher than PU and PEOU at alpha testing (ie, later version hypothesized to be more useful and easier to use than the earlier version). Similar analyses were performed comparing the alpha and beta testing groups on perceived ease of learning and satisfaction.

## Results

### Overview

Qualitative data from individual interviews with youth and providers were reviewed, and content was organized into four main themes: app content, user experience, app use in combination with outpatient therapy, and suggested app modifications. Specific examples quoted from participant interviews are presented in Table 1 to illustrate each of these categories as well as subordinate themes, which are summarized here.



XSL•FO
RenderX

**Table 1.** Common themes and illustrative quotations from formative interviews with youths and providers during alpha and beta testing demonstration sessions.

| Domain and common themes | Participation interview quotes |
|---|---|
| **App content** | |
| Valuable information for teens | • "Having a structured activity rather than just asking a teen what they think their triggers are would be really helpful because sometimes if teens are just asked, they won't know but with the examples, they can begin to recognize things." [Alpha provider]<br>• "It's good that it teaches real facts. Most therapists, the first things they will say about them is that you should not do them, but the app actually leads you to understand why. Most of the time, a therapist will tell you that this is a very bad drug, but the app actually gives you straight facts about the drug and the effects and why it is bad rather than just saying drugs are bad and don't do it. This educates you." [Beta teen] |
| Useful embedded resources | • "I felt like the menu was pretty straightforward and liked that there was a resource section, that way if a kid is having a bad day, they can easily find someone to reach out to." [Beta provider] |
| Variety of activities | • "I think it's good to have a variety of activities as there are different times when different ones are more suitable." [Beta teen] |
| Familiarity | • "It helps since teens are already comfortable with smart phones, it is something familiar whereas treatment may feel less familiar. Teens work better on their phones." [Beta provider] |
| Encourages openness and honesty | • "It is useful for learning more personally...it will make them feel more comfortable answering truthfully and will be easier to answer in an app than in person, especially if you were anti-social or nervous." [Beta teen]<br>• "App could serve as a bridge in circumstances where the teen is thinking it, but just struggling to put it into words or to say out loud. The app might also help teen feel like there is less judgment or would help them answer more honestly, particularly in circumstances where there is a parent in the room." [Beta provider] |
| User-friendly | • "I think it's extremely easy to navigate, very easy. Everything is laid out, there's a list of activities, I knew immediately what to do." [Beta teen]<br>• "I like that it breaks things down so you can select what is most fitting to the circumstance (cravings, resources), that makes it quicker. It's easy to navigate through." [Beta provider] |
| Privacy or restricted access | • "My only concern is parents accessing things that are private. A log in would help or facial recognition. Would want therapist to have access, but parents in general should not because you go to therapy because you can't talk to your parents about things." [Beta teen]<br>• "An option of a password would be a really cool feature for someone who is just trying to get help in secrecy." [Alpha teen] |
| **App use in combination with outpatient therapy** | |
| Improving treatment engagement | • "I would like to use this in session. It would be good to start conversations with my therapist." [Alpha teen] |
| Activities outside of therapy sessions | • "I think it would make a good sort of homework pre-test type thing." [Alpha provider]<br>• "I think it would be great to give homework to clients." [Alpha provider] |
| **App modifications** | |
| More detailed instructions or feedback | • "I'm so used to having my assignments with clear directions. I would like to see that here. I still don't get what to do." [Alpha provider]<br>• "I like that it gives immediate feedback and that it gives further information about the correct answer." [Beta provider] |
| Simplified language | • "I don't talk to my kid patients like this, cognitive distortions, etc. I would probably make this more at a high school reading level, like the words." [Alpha provider] |

## App Content

Providers and teens favored the use of a variety of activities and resources to provide education on topics related to substance use and mental health. Both youth and providers reported that app content was informative, developmentally appropriate, and relevant to adolescents. They found the information to be helpful regarding educating youth on various issues they could encounter and considered the app content as a healthy method for learning and coping. Many providers speculated that youth would find the information more credible than information presented by parents, other authority figures, or the internet. In addition, providers appreciated that content expanded beyond mental health and substance use by including content and activities that addressed family communication and healthy relationships.



XSL•FO
RenderX

Most participants perceived the embedded resources as useful and relevant to the purpose of the app. Providers and teens stated that having in-app access to nationally available resources is valuable for teens to locate immediate services or additional information. In addition, several participants expressed enthusiasm for the app's inclusion of several different types of interactive modules to present information, which they perceived to be more engaging for youth than standard approaches to education in clinical settings (eg, pamphlets and explanations). Participants reported particular excitement about activities such as the *What Would You Do?* scenarios and image maps where users can click on different areas of the screen to reveal educational text or images. Several providers commented that youth might be attracted to game-like features, such as the interactive modules included in the current app.

### User Experience

Providers and youth participants noted their preference for using technology-based tools over paper documentation because of familiarity, user-friendliness, and security of apps. Several teens also recommended that app content be available to youth for self-directed use so that they could choose which activities to complete or view regardless of what may be assigned in a therapy session. A common theme across all participants was the preference for using apps instead of paper documentation to complete questionnaires and assigned activities. The participants cited convenience and accessibility of mobile apps as reasons for preferring apps. In addition, participants valued receiving immediate results and feedback following the completion of the activities.

Participants noted that youths' familiarity with smartphones contributed to their willingness to use and interact with the app. The digital or app-based versions of certain activities were familiar and viewed as reasonable alternatives to standard care for both youth and providers. For instance, youth participants noted how the flash card–style activities mirrored physical cards a therapist may use with the added benefit of always being available, including outside of session.

Many youth participants reported that they would feel more comfortable disclosing personal information on the app than in therapy sessions. Several providers proposed that using an app in addition to verbal discussions in a session could encourage adolescent patients to be more open and candid in sharing information about their thoughts, feelings, and behaviors. Participants emphasized the importance of app privacy and suggested that there be well-defined limits on what user information is released to providers or parents. Several participants noted that using a password log-in would make teenagers feel more comfortable disclosing personal information.

Finally, the provider and teenager participants indicated that they found the app to be user-friendly and easy to navigate. Some specific app features identified as especially helpful in this regard were the search bar for navigation, embedded resources, and extensive feedback for user responses throughout the app. Many participants indicated that these features were necessary if the goal is to have youth use the app regularly.

### App Use in Combination With Outpatient Therapy

Both youth and provider participants endorsed the practicality and appropriateness of using the app in combination with outpatient treatment. They discussed ways to incorporate the app into therapy sessions (eg, to guide in-session discussions and for in-session activities) and outside of therapy as homework activities.

Several providers indicated that the app would help encourage youth patients to identify and clarify their feelings, which may prompt more productive conversations in therapy sessions and improve the youth–provider therapeutic alliance. Most youths shared this viewpoint and commented that they would feel more comfortable entering their mood or feelings into the app to collect their thoughts before discussing with their providers in therapy sessions. In addition, both youths and providers noted that the app would provide a consistent mechanism for completing treatment-related tasks in and out of sessions in a user-friendly format.

Youths and providers indicated that the app could be used for assignments to complete outside of the clinic to reinforce or assess what they learned during therapy sessions. Many youths stated they would use the app several times per week to review content and explore new activities, even if they were not assigned any homework tasks by their providers.

### App Modifications

Participants in the alpha testing phase provided suggestions for modifications to the app. Some of these recommendations were incorporated into the beta version of the app when feasible within the scope of this project. Modifications included improving the overall quality of the content and optimizing the user experience within the app. Participants in the beta testing phase also offered suggestions on improving the future version of the app.

Participants recommended using more extensive instructions and feedback for activities and generally streamlining the app functionality. Before beta testing began, detailed instruction and feedback pages were added to the activities. These additions proved to be effective as beta participants commented on easy navigation, clear instructions, and comprehensive feedback screens. Providers also preferred the simplified language in the app, allowing activities to be more approachable for youths. The simplified text incorporated more definitions throughout the app with the goal of reducing confusion and promoting understanding of key concepts.

Following demonstrations during alpha and beta testing, youths and providers reported on PU, ease of use, ease of learning, and overall satisfaction. Table 2 summarizes the scores from each cohort at each phase of the study, along with comparisons from alpha to beta testing. A significant difference was observed from alpha to beta testing in the PU of the app, whereby providers reported a slightly lower PU at beta testing relative to alpha testing. No other significant differences were observed.



XSL•FO
RenderX

**Table 2.** Youth- and provider-reported perceptions of the app at alpha and beta testing demonstration sessions (N=40).

| Perceptions | Providers | | | | Youths | | | |
|---|---|---|---|---|---|---|---|---|
| | Alpha (n=10), mean (SD) | Beta (n=10), mean (SD) | Mann–Whitney *U* test | *P* value | Alpha (n=10), mean (SD) | Beta (n=10), mean (SD) | Mann–Whitney *U* test | *P* value |
| Perceived usefulness | 2.50 (0.26) | 2.10 (0.42) | 20.00 | .02 | 2.42 (0.41) | 2.33 (0.29) | 41.50 | .53 |
| Perceived ease of use | 2.36 (0.37) | 2.34 (0.36) | 45.00 | .74 | 2.47 (0.56) | 2.41 (0.31) | 42.00 | .58 |
| Perceived ease of learning | 2.50 (0.44) | 2.45 (0.44) | 48.00 | .91 | 2.55 (0.75) | 2.43 (0.41) | 35.50 | .28 |
| Satisfaction | 2.49 (0.28) | 2.28 (0.39) | 30.50 | .14 | 2.55 (0.40) | 2.23 (0.39) | 29.00 | .12 |

# Discussion

## Principal Findings

Over the past decade, there has been a proliferation in mental health apps meant to help people with behavioral health concerns manage their symptoms through coping skills, mood monitoring, and other strategies, sometimes in conjunction with formal therapy, other times as a self-directed, stand-alone app [37]. Research involving such apps has also gained considerable momentum; however, there are few apps that have been developed specifically for youth with substance use using an iterative, user-guided design process, and evidence for such tools remains sparse [38,39]. To address this gap in the clinical toolkit available to clinicians caring for youths with co-occurring substance use and mental health disorders, this mixed methods study involved user-centered design and the development of a web-based mHealth app meant to augment outpatient treatment. Specifically, this study involved an iterative process of evaluation whereby youths with behavioral health treatment experience and providers gave feedback via surveys and interviews following a live demonstration of a new mobile app.

The results indicated that both youths and providers favored the use of an mHealth app over conventional paper documentation because of several advantages of apps, including convenience, accessibility, and capacity to deliver immediate results and feedback even outside of the session. The app content was perceived to be relevant, valuable, and appropriate for the target user population of youths with behavioral health disorders. In addition, youths described feeling more comfortable disclosing personal information within the app as compared with in-person therapy. Providers noted that the use of the app alongside outpatient therapy could be especially helpful if youths complete activities outside of sessions, which would make it easier to discuss difficult topics in session or that could reinforce lessons or skills taught in session, thereby increasing the productivity and impact of treatment.

The evaluation plan was guided by the TAM, which emphasizes the importance of PU and ease of use for promoting the feasibility and acceptability of new technologies and systems. Here, ratings from both youths and providers indicated generally high (positive) ratings for both PU and PEOU, reflecting both appropriate content as well as layout and functionality. Counter to the hypotheses, there were no increased ratings of PU or PEOU from alpha testing to beta testing despite the incorporation of new content and functionality between those sessions. This is likely because of the fact that different participants were recruited for each phase (ie, beta test participants had never seen the alpha version of the app; alpha test participants were not asked to evaluate the beta version of the app). However, there was a significant difference in PU among providers, whereby beta test providers rated the app as somewhat less useful than the alpha test providers. A careful review of specific responses revealed that some beta test providers rated novelty and PU as lower because of the influx of mental health apps on the market, some of which had functionality that is not currently included in the app under investigation here (Bright Path). For instance, some providers noted that other apps include guided meditations and other recorded content to encourage coping skills [40,41], which are not currently part of the Bright Path. As clinician participants had used apps with those sorts of features with patients to enhance the positive effect of their practice, they suggested that the absence of similar features in the current iteration of Bright Path limited its potential usefulness. However, PU was rated favorably, and the critiques raised by the providers in the beta testing phase may be addressed through future software development in future versions of the app.

Participants expressed positive reactions to the variety of activity types and content presented in the app, which were derived from existing cognitive behavioral strategies for addressing substance use and other mental health disorder symptoms. Effective psychotherapy models for SUDs often include elements of motivational interviewing (MI) [42]. MI-related content was not explicitly included in the evaluated versions of the app in this study. Future versions may include more MI-consistent activities. Alternatively, the app may be deployed alongside live psychotherapy (whether in person or telehealth) where the clinician uses MI techniques in session and uses the app to deliver or augment CBT content. This approach has been used in other mHealth studies [43-45]. For instance, young adults who used an app designed to reduce cannabis use reported decreased cravings and cannabis use when delivered alongside 2 motivational enhancement therapy sessions [46]. A similar approach that leverages ecological momentary assessment and remote intervention is currently being studied with youths who exhibit dual disorders [47,48]. This work highlights the importance of considering not only *what* an app includes (content and functionality) but also *how* it is implemented with intended users.

## Limitations

The results of this study should be interpreted in light of these limitations. Although the study involved iterative testing with intended end users of the developed app, the samples of youths



XSL•FO
RenderX

and providers were small at each time point, and different participants were recruited for the alpha and beta tests. Although this prevented within-subject comparisons from alpha to beta testing, which may have resulted in greater ability to detect changes in usability-related variables following edits to the app, it is still valuable to gather fresh perspectives from the intended user population. In addition, participants were recruited locally where the research was conducted, which was a large metropolitan area in the Midwestern United States (population: approximately 2 million people, including the surrounding rural and suburban communities). Perspectives on the usefulness and utility of the app may vary in different geographic regions. For instance, the relative availability of behavioral health specialists may affect the degree to which clinicians would view an app such as this as helpful in as much it might afford more opportunities for self-paced, asynchronous clinical activities so that clinicians could see more patients. Notably, the area where this work took place is a federally designated child mental health shortage area, as is much of the United States (Health Resources and Services Administration, 2021) [49], suggesting that views of youths and providers may have relevance beyond the specific community where the work was conducted. Finally, there was limited demographic diversity among clinician participants (all were women). Together, these factors may have limited the generalizability of the findings to the broader population of adolescents with substance use and mental health disorders and clinicians who provide services to youths with behavioral health disorders. A related limitation was that details about the specific SUDs and mental health disorders that the youths met the diagnostic criteria for were not collected in this study. It is possible that youths with different patterns of mental health symptoms or substance use would view or respond differently to the app. In contrast, the app was not designed to be specific to a particular substance or diagnosis and was intended to be broadly appealing and relevant to youths with SUDs and mental health comorbidities. As such, the generally positive responses and constructive feedback on actionable improvements to the app suggest that this line of research should be maintained with efforts to test future iterations of the app in a more diverse population of youths and providers.

## Conclusions

Mobile apps appear to be a promising, scalable option for supplementing existing therapeutic interventions with adolescents exhibiting co-occurring SUDs and mental health concerns. Apps can provide tools and resources to patients, permitting them to practice what they are learning in therapy in their natural environment between sessions. On the basis of the current results, youths and providers have favorable views about the potential usefulness of apps alongside outpatient therapy, including the specific app tested here, and value having a variety of developmentally appropriate activities geared toward improving treatment engagement in and out of session. Future work should continue to refine the content and functionality, as well as the implementation strategy, for maximal impact in diverse patient populations and clinical settings while also evaluating the efficacy of such tools in improving outcomes for youths with behavioral health disorders.

## Acknowledgments

The authors would like to acknowledge the software development teams at Health Information Technology Solutions and at the Medical University of South Carolina for their contributions to this project, as well as colleagues who provided constructive feedback on the project and this manuscript. This work was supported by the National Institutes of Health/National Institute on Drug Abuse grant K23 DA038257 awarded to ZWA.

## Conflicts of Interest

None declared.

## References

1. Key Substance Use and Mental Health Indicators in the United States: results from the 2018 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. 2019. URL: https://tinyurl.com/n595zzum [accessed 2021-12-12]
2. Trends in annual prevalence of use of various drugs. Monitoring the Future Survey. URL: http://www.monitoringthefuture.org/data/19data/19drtbl2.pdf [accessed 2021-10-12]
3. Robinson ZD, Riggs PD. Cooccurring psychiatric and substance use disorders. Child Adolesc Psychiatr Clin N Am 2016 Oct;25(4):713-722. [doi: 10.1016/j.chc.2016.05.005]
4. Armstrong TD, Costello EJ. Community studies on adolescent substance use, abuse, or dependence and psychiatric comorbidity. J Consult Clin Psychol 2002 Dec;70(6):1224-1239. [doi: 10.1037//0022-006x.70.6.1224] [Medline: 12472299]
5. Kaminer Y, Connor DF, Curry JF. Comorbid adolescent substance use and major depressive disorders: a review. Psychiatry (Edgmont) 2007 Dec;4(12):32-43 [FREE Full text] [Medline: 20436762]
6. Drug use among youth: facts and statistics. National Center for Drug Abuse. URL: https://drugabusestatistics.org/teen-drug-use/ [accessed 2021-10-12]
7. Heradstveit O, Skogen JC, Hetland J, Stewart R, Hysing M. Psychiatric diagnoses differ considerably in their associations with alcohol/drug-related problems among adolescents. A Norwegian population-based survey linked with national patient registry data. Front Psychol 2019 May 8;10:1003 [FREE Full text] [doi: 10.3389/fpsyg.2019.01003] [Medline: 31133937]



XSL•FO
RenderX

8. Kelly TM, Daley DC. Integrated treatment of substance use and psychiatric disorders. Soc Work Public Health 2013;28(3-4):388-406 [FREE Full text] [doi: 10.1080/19371918.2013.774673] [Medline: 23731427]
9. Treatment Episode Data Set (TEDS) discharges from substance abuse treatment services. Substance Abuse and Mental Health Services Administration. URL: https://wwwdasis.samhsa.gov/dasis2/teds_pubs/2011_teds_rpt_d.pdf [accessed 2021-12-12]
10. Henderson JL, Chaim G, Brownlie EB. Collaborating with community-based services to promote evidence-based practice: process description of a national initiative to improve services for youth with mental health and substance use problems. Psychol Serv 2017 Aug;14(3):361-372. [doi: 10.1037/ser0000145] [Medline: 28805421]
11. Brewer S, Godley MD, Hulvershorn LA. Treating mental health and substance use disorders in adolescents: what is on the menu? Curr Psychiatry Rep 2017 Jan;19(1):5. [doi: 10.1007/s11920-017-0755-0] [Medline: 28120255]
12. Shams F, Wong JS, Nikoo M, Outadi A, Moazen-Zadeh E, Kamel MM, et al. Understanding ehealth cognitive behavioral therapy targeting substance use: realist review. J Med Internet Res 2021 Jan 21;23(1):e20557 [FREE Full text] [doi: 10.2196/20557] [Medline: 33475520]
13. Vujanovic AA, Meyer TD, Heads AM, Stotts AL, Villarreal YR, Schmitz JM. Cognitive-behavioral therapies for depression and substance use disorders: an overview of traditional, third-wave, and transdiagnostic approaches. Am J Drug Alcohol Abuse 2017 Jul;43(4):402-415. [doi: 10.1080/00952990.2016.1199697] [Medline: 27494547]
14. Carroll KM, Kiluk BD. Cognitive behavioral interventions for alcohol and drug use disorders: through the stage model and back again. Psychol Addict Behav 2017 Dec;31(8):847-861 [FREE Full text] [doi: 10.1037/adb0000311] [Medline: 28857574]
15. Kazantzis N, Brownfield NR, Mosely L, Usatoff AS, Flighty AJ. Homework in cognitive behavioral therapy: a systematic review of adherence assessment in anxiety and depression (2011-2016). Psychiatr Clin North Am 2017 Dec;40(4):625-639. [doi: 10.1016/j.psc.2017.08.001] [Medline: 29080590]
16. Darnell D, O'Connor S, Wagner A, Russo J, Wang J, Ingraham L, et al. Enhancing the reach of cognitive-behavioral therapy targeting posttraumatic stress in acute care medical settings. Psychiatr Serv 2017 Mar 01;68(3):258-263. [doi: 10.1176/appi.ps.201500458] [Medline: 27745536]
17. Bunnell BE, Nemeth LS, Lenert LA, Kazantzis N, Deblinger E, Higgins KA, et al. Barriers associated with the implementation of homework in youth mental health treatment and potential mobile health solutions. Cognit Ther Res 2021 Apr;45(2):272-286. [doi: 10.1007/s10608-020-10090-8] [Medline: 34108776]
18. Gaynor ST, Lawrence PS, Nelson-Gray RO. Measuring homework compliance in cognitive-behavioral therapy for adolescent depression: review, preliminary findings, and implications for theory and practice. Behav Modif 2006 Sep;30(5):647-672. [doi: 10.1177/0145445504272979] [Medline: 16894234]
19. Wyatt TH, Bayless AK, Krauskopf P, Gaylord N. Using mHealth applications to promote self-managed health behaviors among teens. J Pediatr Nurs 2021;59:164-172. [doi: 10.1016/j.pedn.2021.04.025] [Medline: 33932646]
20. Marsch LA. Leveraging technology to enhance addiction treatment and recovery. J Addict Dis 2012;31(3):313-318 [FREE Full text] [doi: 10.1080/10550887.2012.694606] [Medline: 22873192]
21. Most US teens who use cellphones do it to pass time, connect with others, learn new things. Pew Research Center. URL: https://tinyurl.com/y5ymvv5w [accessed 2021-12-12]
22. Curtis BL, Ashford RD, Magnuson KI, Ryan-Pettes SR. Comparison of smartphone ownership, social media use, and willingness to use digital interventions between generation z and millennials in the treatment of substance use: cross-sectional questionnaire study. J Med Internet Res 2019 Apr 17;21(4):e13050 [FREE Full text] [doi: 10.2196/13050] [Medline: 30994464]
23. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
24. Hutton A, Prichard I, Whitehead D, Thomas S, Rubin M, Sloand E, et al. mHealth interventions to reduce alcohol use in young people: a systematic review of the literature. Compr Child Adolesc Nurs 2020 Sep;43(3):171-202 [FREE Full text] [doi: 10.1080/24694193.2019.1616008] [Medline: 31192698]
25. Mason M, Ola B, Zaharakis N, Zhang J. Text messaging interventions for adolescent and young adult substance use: a meta-analysis. Prev Sci 2015 Feb;16(2):181-188. [doi: 10.1007/s11121-014-0498-7] [Medline: 24930386]
26. Marsch LA, Borodovsky JT. Technology-based interventions for preventing and treating substance use among youth. Child Adolesc Psychiatr Clin N Am 2016 Oct;25(4):755-768 [FREE Full text] [doi: 10.1016/j.chc.2016.06.005] [Medline: 27613350]
27. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. JAMA Pediatr 2017 May 01;171(5):461-469 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
28. ISO 9241-210:2019 Ergonomics of human-system interaction — Part 210: human-centred design for interactive systems. International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2021-12-12]




XSL•FO
RenderX

29. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
30. Mobile Health: Sensors, Analytic Methods, and Applications. Cham: Springer; 2017.
31. Garrido S, Cheers D, Boydell K, Nguyen QV, Schubert E, Dunne L, et al. Young people's response to six smartphone apps for anxiety and depression: focus group study. JMIR Ment Health 2019 Oct 02;6(10):e14385 [FREE Full text] [doi: 10.2196/14385] [Medline: 31579023]
32. Knight JR, Sherritt L, Shrier LA, Harris SK, Chang G. Validity of the CRAFFT substance abuse screening test among adolescent clinic patients. Arch Pediatr Adolesc Med 2002 Jun;156(6):607-614. [doi: 10.1001/archpedi.156.6.607] [Medline: 12038895]
33. Mitchell SG, Kelly SM, Gryczynski J, Myers CP, O'Grady KE, Kirk AS, et al. The CRAFFT cut-points and DSM-5 criteria for alcohol and other drugs: a reevaluation and reexamination. Subst Abus 2014;35(4):376-380 [FREE Full text] [doi: 10.1080/08897077.2014.936992] [Medline: 25036144]
34. Shenoi RP, Linakis JG, Bromberg JR, Casper TC, Richards R, Mello MJ, PEDIATRIC EMERGENCY CARE APPLIED RESEARCH NETWORK. Predictive validity of the CRAFFT for substance use disorder. Pediatrics 2019 Aug;144(2):e20183415 [FREE Full text] [doi: 10.1542/peds.2018-3415] [Medline: 31341007]
35. A technology acceptance model for empirically testing new end-user information systems: theory and results. Massachusetts Institute of Technology. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2021-12-12]
36. Solomon P. The think aloud method: a practical guide to modelling cognitive processes. Inf Process Manag 1995 Nov;31(6):906-907. [doi: 10.1016/0306-4573(95)90031-4]
37. Jaspers MW, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
38. Xu W, Liu Y. mHealthApps: a repository and database of mobile health apps. JMIR Mhealth Uhealth 2015 Mar 18;3(1):e28 [FREE Full text] [doi: 10.2196/mhealth.4026] [Medline: 25786060]
39. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
40. Nunes A, Castro SL, Limpo T. A review of mindfulness-based apps for children. Mindfulness 2020 May 23;11(9):2089-2101. [doi: 10.1007/s12671-020-01410-w]
41. Hutchinson E. Mindfulness apps. School Librarian 2020;68(2):83 [FREE Full text]
42. Help me find an app. One Mind Psyber Guide. URL: https://onemindpsyberguide.org/apps/ [accessed 2021-12-12]
43. Adam A, Jain A, Pletnikova A, Bagga R, Vita A, N Richey L, et al. Use of a mobile app to augment psychotherapy in a community psychiatric clinic: feasibility and fidelity trial. JMIR Form Res 2020 Jul 03;4(7):e17722 [FREE Full text] [doi: 10.2196/17722] [Medline: 32618572]
44. Ben-Zeev D, Drake RE, Corrigan PW, Rotondi AJ, Nilsen W, Depp C. Using contemporary technologies in the assessment and treatment of serious mental illness. Am J Psychiatric Rehab 2012 Oct;15(4):357-376. [doi: 10.1080/15487768.2012.733295]
45. Wilansky P, Eklund JM, Milner T, Kreindler D, Cheung A, Kovacs T, et al. Cognitive behavior therapy for anxious and depressed youth: improving homework adherence through mobile technology. JMIR Res Protoc 2016 Nov 10;5(4):e209 [FREE Full text] [doi: 10.2196/resprot.5841] [Medline: 27833071]
46. Hogue A, Henderson CE, Becker SJ, Knight DK. Evidence base on outpatient behavioral treatments for adolescent substance use, 2014-2017: outcomes, treatment delivery, and promising horizons. J Clin Child Adolesc Psychol 2018;47(4):499-526 [FREE Full text] [doi: 10.1080/15374416.2018.1466307] [Medline: 29893607]
47. Pennou A, Lecomte T, Potvin S, Khazaal Y. Mobile intervention for individuals with psychosis, dual disorders, and their common comorbidities: a literature review. Front Psychiatry 2019;10:302 [FREE Full text] [doi: 10.3389/fpsyt.2019.00302] [Medline: 31130884]
48. Benarous X, Edel Y, Consoli A, Brunelle J, Etter J, Cohen D, et al. Ecological momentary assessment and smartphone application intervention in adolescents with substance use and comorbid severe psychiatric disorders: study protocol. Front Psychiatry 2016 Sep 20;7:157 [FREE Full text] [doi: 10.3389/fpsyt.2016.00157] [Medline: 27703435]
49. Shortage areas. Human Resource & Services Administration. URL: https://data.hrsa.gov/topics/health-workforce/shortage-areas [accessed 2021-12-12]

## Abbreviations

**CBT:** cognitive behavioral therapy
**mHealth:** mobile health
**MI:** motivational interviewing
**PEOU:** perceived ease of use
**PU:** perceived usefulness



XSL•FO
RenderX

**SUD:** substance use disorder
**TAM:** technology acceptance model

*Edited by G Eysenbach; submitted 07.05.21; peer-reviewed by C Shin, M Rampioni; comments to author 11.10.21; revised version received 05.11.21; accepted 18.11.21; published 24.12.21.*

*Please cite as:*
*Adams Z, Grant M, Hupp S, Scott T, Feagans A, Phillips ML, Bixler K, Nallam PT, La Putt D*
*Acceptability of an mHealth App for Youth With Substance Use and Mental Health Needs: Iterative, Mixed Methods Design*
*JMIR Form Res 2021;5(12):e30268*
*URL: https://formative.jmir.org/2021/12/e30268*
*doi:10.2196/30268*
*PMID:34951593*

©Zachary Adams, Miyah Grant, Samantha Hupp, Taylor Scott, Amanda Feagans, Meredith Lois Phillips, Kristina Bixler, Phani Teja Nallam, Dorothy La Putt. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# A Pragmatic Intervention Using Financial Incentives for Pregnancy Weight Management: Feasibility Randomized Controlled Trial

Rebecca Krukowski[1,2], PhD; Brandi Johnson[1], MPH; Hyeonju Kim[1], PhD; Saunak Sen[1], PhD; Riad Homsi[3], MD

[1]Department of Preventive Medicine, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, United States

[2]Department of Public Health Science, University of Virginia, Charlottesville, VA, United States

[3]Just For Women Obstetric Clinic, Memphis, TN, United States

**Corresponding Author:**
Rebecca Krukowski, PhD
Department of Public Health Science
University of Virginia
PO Box 800765
Charlottesville, VA, 22908-0765
United States
Phone: 1 434 924 1000
Email: bkrukowski@virginia.edu

## Abstract

**Background:** Excessive gestational weight gain (GWG) is common and can result in maternal and child health complications. Pragmatic behavioral interventions that can be incorporated into standard obstetric care are needed, and financial incentives are a promising approach.

**Objective:** The aim of this study is to evaluate the feasibility of recruitment, randomization, and retention, as well as treatment engagement and intervention satisfaction, in a behavioral program. The program provided small incentives for meeting behavioral goals of self-weighing and physical activity as well as larger outcome incentives for meeting GWG goals.

**Methods:** We recruited 40 adult women in their first trimester of pregnancy from February 2019 to September 2019 at an obstetric clinic. Participants were randomized to 3 intervention components using a 2×2×2 factorial design: daily incentives for self-weighing (lottery vs certain loss), incentives for adhering to the Institute of Medicine's GWG guidelines based on BMI category (monthly vs overall), and incentives for reaching physical activity goals (yes vs no). Participants were asked to complete daily weigh-ins using the Withings Body wireless scale provided by the study, as well as wear a physical activity tracker (Fitbit Flex 2). Feasibility outcomes of recruitment, randomization, and retention, as well as treatment engagement and intervention satisfaction, were assessed. Weight assessments were conducted at baseline, 32-week gestation, and 36-week gestation.

**Results:** Participants were enrolled at, on average, 9.6 (SD 1.8) weeks' gestation. Of the 39 participants who were oriented to their condition and received the intervention, 24 (62%) were Black or African American, 30 (77%) were not married, and 29 (74%) had an annual household income of less than US $50,000. Of the 39 participants, 35 (90%) completed the follow-up data collection visit. Participants were generally quite positive about the intervention components, with a particular emphasis on the helpfulness of, and the enjoyment of using, the e-scale in both the quantitative and qualitative feedback. Participants who received the loss incentive, on average, had 2.86 times as many days of self-weighing as those who received the lottery incentive. Participants had a relatively low level of activity, with no difference between those who received a physical activity incentive and those who did not.

**Conclusions:** A financial incentive–based pragmatic intervention was feasible and acceptable for pregnant women for promoting self-weighing, physical activity, and healthy GWG. Participants were successfully recruited early in their first trimester of pregnancy and retained for follow-up data collection in the third trimester. Participants demonstrated promising engagement in self-weighing, particularly with loss-based incentives, and reported finding the self-weighing especially helpful. This study supports further investigation of pragmatic, clinic-based financial incentive–based interventions for healthy GWG behaviors.

**Trial Registration:** ClinicalTrials.gov NCT03834194; https://clinicaltrials.gov/ct2/show/NCT03834194

***(JMIR Form Res 2021;5(12):e30578)*** doi:10.2196/30578



XSL•FO
RenderX

**KEYWORDS**

pregnancy; weight; physical activity; self-weighing

# Introduction

## Background

Excessive gestational weight gain (GWG) increases risk for high-cost obstetric conditions such as labor complications and gestational diabetes mellitus for mothers in the short term [1-4]. Excessive GWG also has long-term risks for maternal weight retention [5-7] and childhood obesity [7-9]. For these reasons, GWG is a serious public health concern, particularly because 38% of the women with normal weight, 62% of the women with overweight, and 56% of the women with obesity exceed the GWG guidelines of the Institute of Medicine (IOM) [10]. Pragmatic behavioral interventions that can be incorporated into standard obstetric care are needed.

Financial incentives have been used alone or in combination with other interventions to improve a variety of health outcomes [11-14], including in weight management [15-19]. In addition, incentives are an effective strategy to facilitate smoking cessation among pregnant women [20,21], and incentives may be more attractive for the circumscribed pregnancy period rather than, for example, weight maintenance, which has a much longer time horizon. Surprisingly, however, only 1 study has examined the effect of financial incentives on meeting GWG recommendations; this study found that providing incentives did not increase adherence to GWG outcomes [22]. However, experts have recommended using incentives to encourage health behaviors rather than outcomes [23].

Behaviors that negatively affect health often involve immediate benefits and delayed costs. For example, *eating for two* provides immediate gratification but may lead to excessive GWG [24]. In contrast, behaviors associated with successful weight management (such as daily self-weighing [25-29], weight goals [30], and exercise goals [25,30-33]) often involve immediate time costs with delayed and uncertain health benefits [24]. Thus, incentivizing more immediate GWG-related behaviors may be better than just incentivizing longer-term outcomes.

## Types of Incentives

Incentives, however, can vary widely in their certainty, format, and frequency of distribution [15]. It is generally thought that incentives are more likely to be effective when they are framed as avoiding losses rather than making gains [34] and when rewards are provided immediately [35]. Others have also recommended using small but frequent incentives because these incentives are more visible than large but infrequent payouts [24,36]. A recent study found support for the loss aversion framework in increasing step goal achievement; a greater proportion of participants achieved the daily step goal when it was framed as a loss compared with when the incentive was framed as a gain or a lottery [37]. Another study found that participants who received a lottery-based incentive for reaching a weight loss goal had greater success than the control group [18]. Recent research has also found enhanced engagement and greater weight loss among individuals who received both behavior- and outcome-focused incentives compared with those who did not receive incentives [16,38]. Thus, it will be important to determine what types of incentives are most promising for encouraging GWG-related behaviors and outcomes.

Consistent with previous research and based on self-regulation theory principles [39], this study examines the impact of small incentives for meeting behavioral goals of self-weighing and physical activity as well as larger outcome incentives for meeting GWG goals. The aim of this pilot study is to evaluate the feasibility of recruitment, randomization, and retention, as well as treatment engagement and intervention satisfaction.

# Methods

## Study Design

This pilot study recruited participants from February 2019 to September 2019 from an obstetric clinic with 1 physician in Memphis, Tennessee. Participants were randomized to 3 intervention components using a 2×2×2 factorial design: (1) daily incentives for self-weighing on a wireless scale (lottery vs certain loss), (2) incentives for adhering to the IOM's GWG guidelines based on BMI category (monthly vs overall), and (3) incentives for reaching physical activity goals (yes vs no). The intervention lasted approximately 6 months, depending on participants' gestational age at baseline. Assessments were conducted at baseline, 32-week gestation, and 36-week gestation (if the participant had not yet delivered). Participants were asked to complete daily weigh-ins using the Withings Body wireless scale provided by the study, as well as wear a physical activity tracker (Fitbit Flex 2; Google LLC). Incentives (US $30 for each follow-up visit) were used to facilitate high retention at both follow-up data collection visits, which took place at the obstetric clinic. This study was registered with ClinicalTrials.gov (NCT03834194) and approved by the institutional review board of the University of Tennessee Health Science Center.

## Sample

Potential participants were identified by the clinic's nurses and obstetrician at their pregnancy confirmation visit or self-identified through recruitment materials posted at the obstetric clinic (including in the examination rooms and waiting room; Multimedia Appendix 1). They were encouraged to meet with the study team in the obstetric office if they wished to learn more about the study and potentially enroll in the study. Interested individuals were evaluated for the eligibility criteria, and electronic informed consent was obtained. The obstetric clinic has approximately 16 pregnancy confirmation visits per month.

To be eligible, individuals needed to be aged at least 18 years and no more than 13 weeks', 0 days' gestation upon recruitment (based on the date of their last menstrual period and then confirmed by their physician at their first prenatal visit) because GWG-focused interventions that begin in the first trimester are more effective [40]. Individuals also needed to verify that they were having a singleton pregnancy by ultrasonography because



XSL•FO
RenderX

of the different GWG guidelines for women with multiple gestation [7], and participants were enrolled only if the physician believed that the intervention would be safe for them. Additional eligibility criteria included (1) having a BMI of ≥18.5 kg/m$^2$ (because of the infrequency of women with underweight) and (2) having access to wireless internet or a Bluetooth-connected device to facilitate data collection for participants' self-weighing behavior and physical activity.

## Enrollment

A total of 40 participants provided informed consent and were enrolled on the web using the Way to Health web platform developed by the University of Pennsylvania, which integrated enrollment, randomization, surveys, automated delivery of study email messages, and transfer of data from the study's wireless devices [41]. All participants received a Fitbit Flex 2 and a Withings Body scale at enrollment to track their physical activity and weight over time. Before randomization, research staff members oriented the participants to the devices (including providing handouts to review if questions arose later). The staff members then assisted participants with setting up Fitbit and Withings accounts, downloading the apps to their smartphones, pairing the devices with Bluetooth, and authorizing the transfer of data to the Way to Health platform. Participants were instructed to open the apps each day to transmit their data to the Way to Health platform.

## Randomization

After enrollment, participants were randomized into 1 of 8 conditions (Table 1), that represented all combinations of the 3 different components. Participants received an intervention orientation message and handout that detailed the components to which they were randomized, the recommendation to weigh daily, and a GWG recommendation tailored to BMI category (with overall or monthly goals). They also received a handout that provided strategies for achieving healthy GWG (Multimedia Appendix 2).

**Table 1.** Randomized conditions (N=40).

| Condition | Self-weighing | Weight goal | Physical activity goal | Participants, n (%) |
|---|---|---|---|---|
| 1 | Lottery | Monthly | No | 5 (13) |
| 2 | Lottery | Overall | Yes | 5 (13) |
| 3 | Lottery | Overall | No | 5 (13) |
| 4 | Lottery | Monthly | Yes | 5 (13) |
| 5 | Loss | Monthly | No | 5 (13) |
| 6 | Loss | Overall | Yes | 5 (13) |
| 7 | Loss | Overall | No | 5 (13) |
| 8 | Loss | Monthly | Yes | 5 (13) |

## Procedures

### *Self-weighing Lottery-Based Incentive*

Participants randomized to a condition with the self-weighing lottery-based incentive (conditions 1, 2, 3, and 4) were asked to pick a *lucky* number from 0 to 99 at randomization. They were informed that there would be a daily lottery for which they would be eligible if they weighed themselves on the previous day on the scale provided by the study team. Then, for each day of their pregnancy, participants were informed of the study's randomly generated winning lottery number. Participants who weighed themselves on the previous day and who had a 1- or 2-digit match between their *lucky* number and the number that was drawn were notified of their reward, consistent with previous research [42]. A 2-digit match (1 in 100 chance) yielded a US $15 incentive and a single-digit match (1 in 5 chance) yielded a US $2 incentive in the form of an Amazon gift card. Participants who did not weigh themselves the previous day were informed that they could have won incentives, consistent with loss aversion principles [43] (see Multimedia Appendix 3 for sample email messages). Participants also received automated email messages when they did not transmit data, reminding them to self-weigh and sync their devices. The daily winnings were accumulated for the week starting Monday and were disbursed the following Monday. The maximum payout for this component was US $112 in total.

### *Self-weighing Loss-Based Incentive*

Participants randomized to a condition with the self-weighing loss-based component (conditions 5, 6, 7, and 8) had a weekly balance of US $3.50 at the beginning of each week in their account. Then, for each day that they did not weigh, US $0.50 was subtracted from this account. Participants who did not weigh themselves the previous day were informed that they had lost US $0.50, consistent with loss aversion principles [43] (see Multimedia Appendix 3 for sample email messages). The daily winnings were accumulated for the week starting Monday and were disbursed the following Monday. The maximum payout for this component was US $112 total.

### *Monthly GWG Goal Incentive*

Participants randomized to a condition with the monthly GWG goal component (conditions 1, 4, 5, and 8) received US $14 per month if their monthly GWG was within the recommended monthly range for their BMI category (Table 2), which they had received in a handout at randomization. Participants received a monthly email regarding whether their GWG was within the recommended range or not and, thus, whether they received the monthly GWG incentive (see Multimedia Appendix 3 for sample email messages).



XSL•FO
RenderX

The monthly GWG goals were constructed based on the IOM's recommended range of total GWG tailored to BMI category, the minimum recommended GWG in the first trimester for all BMI categories (1.1 lb), and the range of recommended weekly GWG in the second and third trimesters by BMI category [7]. Participants were eligible for the incentive starting from the first full month of pregnancy after the start of their participation. To receive the incentive, the participant needed to self-weigh on the study-provided scale on at least 1 day in the last week of the gestational month as well as have at least one weight in the previous month to calculate GWG within the month.

**Table 2.** Monthly weight gain goals by weight status[a].

| | Monthly weight gain goals (lb) | | |
|---|---|---|---|
| | Normal weight | Overweight | Obese |
| **First trimester** | | | |
| Weeks 5-8 | <2.2 | <2.2 | <2.2 |
| Weeks 9-12 | <2.2 | <2.2 | <2.2 |
| **Second trimester** | | | |
| Weeks 13-16 | 2.8-5 | 1.4-3.6 | 0.8-2.8 |
| Weeks 17-20 | 2.8-5 | 1.4-3.6 | 0.8-2.8 |
| Weeks 21-24 | 2.8-5 | 1.4-3.6 | 0.8-2.8 |
| **Third trimester** | | | |
| Weeks 25-28 | 2.8-5 | 1.4-3.6 | 0.8-2.8 |
| Weeks 29-32 | 2.8-5 | 1.4-3.6 | 0.8-2.8 |
| Weeks 33-36 | 2.8-5 | 1.4-3.6 | 0.8-2.8 |

[a]There was no lower bound for the first trimester weight goal so as to not penalize pregnant women who have trouble gaining weight because of nausea.

### *Overall GWG Goal Incentive*

Participants randomized to a condition with the overall GWG goal component (conditions 2, 3, 6, and 7) were provided the IOM's overall GWG recommendation based on their BMI at randomization (Table 3). The overall GWG goals were constructed based on the IOM's recommended range of total GWG and the mean of recommended weekly GWG in the second and third trimesters. These values were adjusted for 8 weeks (for data collection at 32 weeks) and 4 weeks (for data collection at 36 weeks) from a full-term pregnancy at 40 weeks. If the participant met the goal, she received the overall GWG incentive after the 36-week visit, unless she delivered before this visit (in which case she received the incentive based on her 32-week visit weight upon notification that she had given birth).

**Table 3.** Overall weight gain goals by weight status (adjusted goals based on data collection at 32 weeks' and 36 weeks' gestation).

| Gestational weeks | Overall weight gain goals (lb) | | |
|---|---|---|---|
| | Normal weight | Overweight | Obese |
| 32 | 17-27 | 10.2-20.2 | 7-16 |
| 36 | 21-31 | 12.6-22.6 | 9-18 |

### *Weekly Physical Activity Goal Incentive*

Participants randomized to an arm with the physical activity incentive component (conditions 2, 4, 6, and 8) were encouraged to achieve a goal of 150 minutes of physical activity per week based on the guideline from the American College of Obstetricians and Gynecologists [44]. Participants received US $3.50 if they met the activity goal each week. *Fairly active minutes* and *very active minutes* based on the Fitbit activity tracker programming counted toward the 150-minute goal [45]. Participants who did not meet the activity goal were notified that they would have received US $3.50 had they met their activity goal (see Multimedia Appendix 3 for sample email messages). The maximum payout for this component was US $105 (US $3.50 per week for a participant who joined the study at 6 weeks' gestation).

## Measures

### *Sociodemographic Characteristics*

At baseline, participants reported their age, race, ethnicity (Hispanic or non-Hispanic), marital status, educational background, employment status, income, and the number of children in their household.

### *Participant Recruitment*

Recruitment yields were calculated based on the number of participants who indicated interest in the study and the number of participants who were randomized.



XSL•FO
RenderX

### *Perceptions About the Effectiveness of the Incentive Conditions*

The baseline assessments included participants' perceptions about which of the incentive types would be more effective (daily self-weighing: lottery vs loss-based; weight gain goals: monthly vs overall).

### *Intervention Satisfaction*

At the 32-week follow-up assessment, participants were asked about their satisfaction with the intervention. Specifically, they were asked to indicate how satisfied they were with each component of the intervention on a 5-point scale from 1 (not at all) to 5 (extremely), including an option to indicate *not applicable* if they had not received the component. They were asked to separately rate the helpfulness and enjoyment aspect of the emails, electronic scales, and Fitbit activity trackers. They were also asked to separately rate the usefulness and enjoyment aspect of the lottery incentive, loss incentive, the monthly weight gain goal, overall weight gain goal, and the weekly physical activity goal. In addition, they were asked to rate how likely it was that they would recommend using an electronic scale, a fitness tracker, setting a goal for GWG, and setting a goal for physical activity for managing GWG to a friend. Finally, at the end of the structured questionnaire, participants were asked to respond to 3 open-ended questions regarding the recommendations they would make for changes to the program, aspects of the program that were most helpful, and anything else that they wanted to share that they thought might help improve pregnancy weight management programs in the future.

### *Intervention Engagement*

Participants' daily self-weighing behavior (coded each day as present or absent) and weekly physical activity (measured in active minutes) were monitored electronically because they were measures of treatment engagement.

### *Retention*

Program retention was observed from the number of participants who consented to participate and then completed the follow-up data collection visits at 32 weeks' and 36 weeks' gestation. Participants were alerted that they would be withdrawn from the study if they exhibited extreme physical activity or restricted weight gain as potential symptoms of an eating disorder. Although the physician was alerted several times to instances of weight loss, no participant was recommended to be withdrawn for this reason.

### *GWG Measurement Visits*

The secondary outcomes included GWG from baseline to the final data collection point before delivery (32 weeks' and 36 weeks' gestation). At all measurement visits, participants' weights were recorded in kilograms on a calibrated research-grade scale in duplicate, with the participants wearing light clothing and no shoes. In addition to the measured weight at baseline, participants also reported their preconception weight. Participants' GWG goals were calculated based on self-reported preconception BMI. There is strong concordance between self-reported preconception weight and measured pregravid weight [46]. The weight at week 36 was the default outcome, except for participants who delivered earlier than 36 weeks, in which case their weight at week 32 was used. This approach is an attempt to have a final observation of GWG for all participants, regardless of whether they delivered earlier than 36 weeks.

## Power Estimates

This is a feasibility study, the primary purpose of which is to gather data that would evaluate the feasibility of recruitment, randomization, and retention, as well as treatment engagement and intervention satisfaction [47]. An additional aim of this study is to pilot the intervention components, consistent with the preparatory stage of the multiphase optimization strategy framework [48]. The trial was not powered to detect a significant difference among the conditions on treatment engagement or GWG.

## Statistical Analysis

We described characteristics of the sample using counts and percentages for categorical data and means and SDs for continuous data. We used generalized linear models with log link function and variance function proportional to the mean for analyzing the 2×2×2 factorial design for 2 outcomes: days of self-weighing and mean of each participant's weekly physical activity minutes. The model included main effects for the 3 intervention components (self-weighing, weight goal, and physical activity goal) and all 2- and 3-factor interactions for constructing 8 conditions in the 2×2×2 factorial analysis. Because of the small sample size and our primary interest in the main effects, the analysis focused on reporting the main effects. The reference levels were loss incentive for self-weighing, overall GWG goal, and no physical activity goal. Participants' GWG values according to the guidelines as an ordinal variable (below, within, and exceeded) were not included in this analysis because of the small sample size. We conducted 2-tailed *t* tests to test the null hypothesis of no difference in the incentive amounts received by component and overall. Minutes of physical activity were calculated based on days for which data were transmitted. All analyses were implemented using R (version 4.0.2; The R Foundation for Statistical Computing).

# Results

## Recruitment and Retention

A total of 41 participants indicated interest in the study over the 6-month recruitment period and completed the informed consent procedure, but 1 (2%) participant refused to participate after completing the informed consent process. In all, 40 participants were randomized; however, 1 (3%) participant was immediately withdrawn because of a staff member's error in assessing this participant's age eligibility (she was aged <18 years). This participant was not oriented to her condition and did not receive any intervention. Of the 39 participants who were oriented to their condition and received the intervention, 35 (90%) completed the follow-up data collection visit at 32 weeks' and 36 weeks' gestation; 1 (3%) participant delivered before the 36-week visit, and her 32-week weight was used. Among the 4 participants who did not complete the follow-up data collection visit, 1 (25%) relocated, 1 (25%) miscarried, 1



XSL•FO
RenderX

(25%) delivered her baby before the 32-week visit, and we were unable to contact 1 (25%) participant.

As can be seen in Table 4, the participants predominantly identified as Black or African American (24/39, 62%) and non-Hispanic (36/39, 92%). Among the 39 participants, for 12 (34%), this was their first pregnancy. Of the 39 participants, 30 (77%) were not married, 27 (70%) had at least some college education, 29 (74%) were employed, and 29 (74%) had an annual household income of less than US $50,000. With regard to managing GWG, of the 39 participants, 32 (82%) believed that the lottery-based daily self-weighing incentive would work better to help them and 31 (80%) believed that the monthly GWG goals would work better to help them. The participants were successfully recruited in the first trimester (mean gestational weeks at enrollment 9.6, SD 1.8), and they participated in the program for between 24 and 30 weeks.



XSL•FO
RenderX

**Table 4.** Sample characteristics (N=39).

| | Values |
|---|---|
| Age at enrollment, mean (SD) | 29.1 (12.5) |
| Gestational weeks at enrollment, mean (SD) | 9.6 (1.8) |
| Number of children in household, mean (SD) | 1.1 (0.9) |
| **Prepregnancy BMI categories, n (%)** | |
| Normal weight (BMI 18.5-24.9 kg/m$^2$) | 17 (44) |
| Overweight (BMI 25-29.9 kg/m$^2$) | 13 (33) |
| Obese (BMI ≥30 kg/m$^2$) | 9 (23) |
| Hispanic ethnicity, n (%) | 3 (8) |
| **Race, n (%)** | |
| White | 10 (26) |
| Black or African American | 24 (62) |
| Unknown or multiple races | 5 (13) |
| **Marital status, n (%)** | |
| Married | 9 (23) |
| Never married | 18 (46) |
| A member of an unmarried couple | 12 (31) |
| **Education attainment, n (%)** | |
| Some high school | 2 (5) |
| High school graduate or GED[a] | 10 (26) |
| Some college or technical school | 20 (51) |
| College, 4 years or more (college graduate) | 7 (18) |
| **Employment status, n (%)** | |
| Employed for wages | 29 (74) |
| Self-employed | 2 (5) |
| Out of work | 2 (5) |
| Does not work outside the home | 1 (3) |
| Student | 4 (10) |
| Missing | 1 (3) |
| **Annual household income (US $) n (%)** | |
| ≤24,999 | 18 (46) |
| 25,000-49,999 | 11 (28) |
| 50,000-74,999 | 4 (10) |
| ≥75,000 | 5 (13) |
| Missing | 1 (3) |
| **Which type of incentive do you think would work better for you to manage gestational weight gain?, n (%)** | |
| **Daily self-weighing** | |
| Lottery | 32 (82) |
| Loss | 7 (18) |
| **Weight gain goals** | |
| Monthly weight gain goal | 31 (80) |
| Overall weight gain goal | 8 (20) |



XSL•FO
RenderX

[a]GED: General Educational Development.

## Intervention Satisfaction

Participants were generally quite positive about the intervention components (Table 5), with a mean score for each component of ≥3.8. Consistently, the mean score for the helpfulness and enjoyment aspect of the electronic scale was slightly higher than that for the other components; this positive view of self-weighing and weight tracking throughout pregnancy was also noted by 80% (28/35) of the participants in qualitative feedback regarding the aspects of the program that were most helpful to them. A participant stated, "Most helpful was having to weigh every day, it kept me in tune with where I was and what I needed to do differently." In addition, in the qualitative responses, several participants indicated that it would have been helpful to have the option to receive SMS text messages rather than emails. A few participants also reported technical challenges, particularly with the e-scale. A participant noted, "I believe this is a great program for those who were already active prior to pregnancy," perhaps reflecting the relatively low-intensity behavioral intervention for physical activity. Finally, participants indicated that additional features that might be helpful for this program would be dietary tracking and group meetings for peer support.

**Table 5.** Summary of intervention satisfaction (N=35).

| | Values, mean (SD) |
|---|---|
| **Helpfulness for managing GWG[a] (1=not at all helpful; 5=extremely helpful)** | |
| Emails | 4.4 (0.9) |
| Electronic scale | 4.5 (0.9) |
| Fitbit | 4.2 (1.0) |
| **Usefulness for managing GWG (1=not at all useful; 5=extremely useful)** | |
| Lottery-based self-weighing incentive | 3.9 (1.1) |
| Loss-based self-weighing incentive | 3.9 (1.0) |
| Monthly weight gain goal | 4.1 (0.9) |
| Overall weight gain goal | 4.0 (0.9) |
| Weekly physical activity incentive | 3.8 (1.1) |
| **How much did you enjoy the following? (1=not at all enjoyable; 5=extremely enjoyable)** | |
| Using electronic scale | 4.6 (0.7) |
| Using Fitbit | 4.0 (1.2) |
| Frequency of receiving incentives | 4.2 (0.8) |
| Frequency of receiving emails | 4.1 (1.2) |
| **How much did you enjoy receiving...incentives? (1=not at all enjoyable; 5=extremely enjoyable)** | |
| Lottery-based self-weighing incentive | 4.3 (1.1) |
| Loss-based self-weighing incentive | 4.3 (1.1) |
| Monthly weight gain goal | 4.3 (1.0) |
| Overall weight gain goal | 4.2 (1.1) |
| Weekly physical activity incentive | 4.1 (1.2) |
| **Would you recommend...for managing GWG to a friend? (1=not at all likely; 5=extremely likely)** | |
| Using an electronic scale | 4.7 (0.7) |
| Using a fitness tracker | 4.5 (0.9) |
| Setting a goal for weight gain | 4.6 (0.6) |
| Setting a goal for physical activity | 4.6 (0.7) |

[a]GWG: gestational weight gain.

## Intervention Adherence

The mean number of days that participants self-weighed on the study scale was 44.5 (SD 57.5) days. Participants who were randomized to receive the loss incentive had a mean of 68.0 (SD 71.4) days of self-weighing compared with 19.8 (SD 18.8) days among those participants who were randomized to the lottery incentive. There was a significant main effect such that the mean difference in log means (or the



XSL•FO
RenderX

log of the ratio of the means) was –1.05. The ratio of the means was 1/exp(1.05) = 1/2.86 = 0.35 (Table 6). In other words, participants who received the loss incentive, on average, had 2.86 times as many days of self-weighing as those who received the lottery incentive, or inversely the mean frequency of self-weighing with the lottery incentive was 0.35 times the loss incentive. Consistent with this greater adherence to self-weighing among those who received the loss incentive, participants in the loss incentive condition earned, on average, US $59.58, in self-weighing incentives in comparison with US $25.66 earned by those in the lottery incentive condition, although the amount that they could have earned in each condition was the same (US $112).

**Table 6.** Factorial main effects for all outcomes (N=39).

| 2×2×2 intervention components | Days of self-weighing | | Mean of participants' weekly PA[a] | |
|---|---|---|---|---|
| | Mean ratio (95% CI) | Estimate (SE)[b] | Mean ratio (95% CI) | Estimate (SE) |
| Self-weighing (reference=loss), lottery | 0.35 (0.15-0.80) | –1.05[c] (0.40) | 5.38 (0.34-85.92) | 1.68 (1.36) |
| Gestational weight gain goal (reference=overall), monthly | 0.69 (0.30-1.58) | –0.36 (0.40) | 0.24 (0.015-3.80) | –1.43 (1.36) |
| PA goal (reference=no incentives), incentives | 1.26 (0.55-2.88) | 0.23 (0.40) | 0.79 (0.05-12.56) | –0.24 (1.36) |

[a]PA: physical activity.

[b]Estimates and SEs are on the logarithmic scale; the mean ratio is obtained by exponentiating the main effect estimate.

[c]$P<.01$.

Participants, on average, had a relatively low level of activity, with a mean number of active minutes of 8.7 (SD 18.5) per week. Participants who were randomized to receive the physical activity incentive had a mean of 5.4 active minutes per week compared with 12.0 active minutes per week among those participants who were not randomized to receive physical activity incentives. However, there was no significant main effect. Of the 39 participants, 2 (5%) achieved the physical activity goal, and they received the US $3.50 incentive in just 1 week each.

Among the 35 participants who completed the follow-up data collection, 7 (20%) had weight gain below the GWG guidelines, 13 (37%) were within the guidelines, and 15 (43%) exceeded the guidelines. A similar proportion of individuals exceeded the guidelines, regardless of whether they were randomized to receive GWG incentives monthly (8/35, 44%) or overall (7/35, 41%). Of the 19 participants who were randomized to the monthly GWG incentive component, 8 (42%) received at least one monthly US $14 incentive (with a range of 1-5 monthly incentives received).

Participants earned, on average, US $43.05 (range US $0 to US $211) in total. Not surprisingly, participants who were within the guidelines earned more on average (US $91.27) compared with those who were below the guidelines (US $16.50) and those who exceeded the guidelines (US $19.87).

## Discussion

### Principal Findings

We demonstrated the feasibility of conducting a randomized controlled trial with a 2×2×2 factorial design at an obstetric clinic that provided financial incentives for self-weighing, physical activity, and GWG within the IOM guidelines. We successfully recruited participants early in their first trimester of pregnancy and retained 90% (35/39) of these participants for follow-up data collection in the third trimester. We were also successful in randomizing participants to this complex randomization scheme and providing the assigned intervention components. In addition, the intervention was well-received, with high satisfaction ratings from the participants for all the components and a particular focus on the helpfulness of self-weighing and weight tracking. The participants engaged in self-weighing approximately 7 days per month, in contrast to previous research indicating that regular self-weighing is uncommon among pregnant women without intervention [49]. However, there was low engagement in physical activity among these participants, regardless of whether they received an incentive for reaching the 150-minute physical activity goal.

On the basis of the projected clinic flow of 16 pregnancy confirmation appointments per month, we were successful in recruiting more than one-third of the patients per month, with more than 6 participants recruited to the study per month. Notably, all but 2 participants who indicated initial interest were eligible and interested in being randomized. In addition, retention was very good, with only 10% (4/39) of the participants not participating in follow-up data collection (including a participant who had a miscarriage and a participant who gave birth before the 32-week visit).

Contrary to previous literature that has reported that participants prefer incentive schemes *other* than lotteries [50-52], the participants in this study indicated that they believed that the lottery-based incentive for self-weighing would be more helpful for them in managing GWG than the loss-based incentive. Surprisingly, contrary to their stated preference, participants who were randomized to receive the loss-based incentive had significantly greater self-weighing engagement compared with those randomized to the lottery-based incentive. This finding is consistent with a recent study, which found that loss-based incentives were more effective in increasing step goal attainment than a lottery-based incentive [37].

Although regular physical activity is recommended for a healthy pregnancy [44], most women in general are physically inactive, particularly during pregnancy [53]. Consistent with these previous findings, we found that women in our sample had a very low level of physical activity overall. The physical activity



XSL•FO
RenderX

incentives seemed to do little to increase physical activity. On the basis of a recent meta-analysis [54], more intensive intervention such as supervised exercise may be necessary for increasing physical activity during pregnancy.

This study is the first to offer both *process* incentives (for behaviors) and *outcome* incentives (for achieving monthly or overall GWG), within the context of gestational weight management. However, there were no clear trends in this study as to whether an incentive for achieving a monthly or an overall GWG goal was associated with gaining weight within the IOM guidelines. A similar proportion of participants in this sample (13/36, 37%) gained weight within the guidelines compared with the national prevalence (32.1%) [10]. Although this study was not powered to detect differences in GWG recommendation adherence, this finding is consistent with previous research indicating that incentives did not increase adherence to GWG recommendations [22]. Should the lack of impact of financial incentives on GWG outcomes be confirmed in future fully powered research, it may be that financial incentives are more powerful at encouraging behaviors associated with GWG than GWG outcomes themselves. In that case, it would then be essential to determine whether the costs to achieve certain behaviors in pregnancy (such as self-weighing) are beneficial independently. For example, it is possible that incentives for self-weighing may not lead to optimal GWG (as a categorical variable), but incentives for self-weighing may reduce the mean amount of GWG [55] and thus reduce the likelihood of negative maternal health outcomes.

This study includes several notable strengths and weaknesses. First, the recruited sample was quite diverse in race, income level, and BMI category and, thus, included many individuals who often are not included in research [56,57]. However, this study had a small sample to determine feasibility and recruited from only 1 obstetric clinic; thus, future fully powered research is necessary. In addition, the incentive strategies and amounts used in this study were only a few of the many possible approaches, and it is possible that there are more effective alternatives that should be tested. A further limitation is that the data reported here and the incentives provided are based on the data received by the research team and not necessarily all of the data sent by the participants. This is an important distinction because 5% (2/39) of the participants reported having technical problems in the program evaluation. It is also possible that the low levels of physical activity were due to undetected difficulties in transmitting physical activity data. In addition, although we did not report all the costs of the intervention, the unreported costs (including the costs of the scale and the Fitbit activity tracker) were the same for all participants.

This feasibility study provides valuable information for future research, including the expected rate of clinic-based recruitment per physician for a low-intensity intervention, the high rate of retention with convenient clinic-based data collection, and expected adherence and costs for a financial incentive–based intervention. In future research, given our few unresolved technical difficulties, it may be important to set up a telephone visit after a few days to ensure that participants are able to use the technology in their home environment and test using multiple modalities (SMS text messages, telephone, and email) for investigating potential technical problems. The findings from this study also indicate enthusiasm for self-weighing among pregnant women as a strategy for gestational weight management, which may be important in designing future studies. These results also may indicate that preferences for an intervention strategy may not translate into the more effective strategy, which has also been demonstrated in other research [58]. Our findings indicate that although the loss-based incentives may not be perceived as more effective at baseline, they are a promising strategy for increasing self-weighing behaviors. Finally, this feasibility study suggests that more intense intervention strategies (such as meal replacements, motivational interviewing or problem-solving sessions, calorie goals, and dietary self-monitoring with feedback from an interventionist) may be necessary for increasing moderate physical activity among pregnant women and adherence to the GWG guidelines [59,60]. For example, future research in this area could be a factorial experiment that examines the combination of outcome and behavioral incentives with more intense intervention strategies.

## Conclusions

This pilot randomized controlled trial indicates that a financial incentive–based pragmatic intervention is feasible and acceptable for pregnant women for promoting self-weighing, physical activity, and healthy GWG. Participants were successfully recruited early in their first trimester of pregnancy and retained for follow-up data collection in the third trimester. Participants demonstrated promising engagement in self-weighing, particularly with loss-based incentives, and reported finding the self-weighing especially helpful. This study supports further investigation of pragmatic, clinic-based interventions for healthy GWG.

## Acknowledgments

The research reported here was supported by the Jane Nuckolls Grant to Fight Hunger Fund.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Recruitment brochure.
[PDF File (Adobe PDF File), 4128 KB - formative_v5i12e30578_app1.pdf ]



XSL•FO
RenderX

Multimedia Appendix 2
Study handout: strategies for achieving healthy gestational weight gain.
[PDF File (Adobe PDF File), 40 KB - formative_v5i12e30578_app2.pdf ]

Multimedia Appendix 3
Example email messages for each intervention condition.
[DOCX File , 14 KB - formative_v5i12e30578_app3.docx ]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1224 KB - formative_v5i12e30578_app4.pdf ]

## References


1. Macdonald-Wallis C, Tilling K, Fraser A, Nelson SM, Lawlor DA. Gestational weight gain as a risk factor for hypertensive disorders of pregnancy. Am J Obstet Gynecol 2013 Oct;209(4):327.e1-327.17 [FREE Full text] [doi: 10.1016/j.ajog.2013.05.042] [Medline: 23711667]
2. Brunner S, Stecher L, Ziebarth S, Nehring I, Rifas-Shiman SL, Sommer C, et al. Excessive gestational weight gain prior to glucose screening and the risk of gestational diabetes: a meta-analysis. Diabetologia 2015 Oct;58(10):2229-2237. [doi: 10.1007/s00125-015-3686-5] [Medline: 26141788]
3. Thorsdottir I, Torfadottir JE, Birgisdottir BE, Geirsson RT. Weight gain in women of normal weight before pregnancy: complications in pregnancy or delivery and birth outcome. Obstet Gynecol 2002 May;99(5 Pt 1):799-806. [doi: 10.1016/s0029-7844(02)01946-4] [Medline: 11978290]
4. Stotland NE, Cheng YW, Hopkins LM, Caughey AB. Gestational weight gain and adverse neonatal outcome among term infants. Obstet Gynecol 2006 Sep;108(3 Pt 1):635-643. [doi: 10.1097/01.AOG.0000228960.16678.bd] [Medline: 16946225]
5. Phelan S. Pregnancy: a "teachable moment" for weight control and obesity prevention. Am J Obstet Gynecol 2010 Feb;202(2):135.e1-135.e8 [FREE Full text] [doi: 10.1016/j.ajog.2009.06.008] [Medline: 19683692]
6. Mamun A, Kinarivala M, O'Callaghan MJ, Williams GM, Najman JM, Callaway LK. Associations of excess weight gain during pregnancy with long-term maternal overweight and obesity: evidence from 21 y postpartum follow-up. Am J Clin Nutr 2010 May;91(5):1336-1341. [doi: 10.3945/ajcn.2009.28950] [Medline: 20237138]
7. Weight Gain During Pregnancy: Reexamining the Guidelines. Washington: National Academies Press; 2009.
8. Hu Z, Tylavsky FA, Han JC, Kocak M, Fowke JH, Davis RL, et al. Maternal metabolic factors during pregnancy predict early childhood growth trajectories and obesity risk: the CANDLE Study. Int J Obes (Lond) 2019 Oct;43(10):1914-1922 [FREE Full text] [doi: 10.1038/s41366-019-0326-z] [Medline: 30705389]
9. Liao X, Yu Y, Marc I, Dubois L, Abdelouahab N, Bouchard L, et al. Prenatal determinants of childhood obesity: a review of risk factors. Can J Physiol Pharmacol 2019 Mar;97(3):147-154. [doi: 10.1139/cjpp-2018-0403] [Medline: 30661367]
10. Deputy NP, Sharma AJ, Kim SY. Gestational Weight Gain - United States, 2012 and 2013. MMWR Morb Mortal Wkly Rep 2015 Nov 06;64(43):1215-1220 [FREE Full text] [doi: 10.15585/mmwr.mm6443a3] [Medline: 26540367]
11. Higgins ST, Budney AJ, Bickel WK, Foerg FE, Donham R, Badger GJ. Incentives improve outcome in outpatient behavioral treatment of cocaine dependence. Arch Gen Psychiatry 1994 Jul;51(7):568-576. [doi: 10.1001/archpsyc.1994.03950070060011] [Medline: 8031230]
12. Pope L, Harvey-Berino J. Burn and earn: a randomized controlled trial incentivizing exercise during fall semester for college first-year students. Prev Med 2013 Mar;56(3-4):197-201. [doi: 10.1016/j.ypmed.2012.12.020] [Medline: 23295170]
13. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
14. Strohacker K, Galarraga O, Williams DM. The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. Ann Behav Med 2014 Aug;48(1):92-99 [FREE Full text] [doi: 10.1007/s12160-013-9577-4] [Medline: 24307474]
15. Jeffery RW. Financial incentives and weight control. Prev Med 2012 Nov;55 Suppl:S61-S67 [FREE Full text] [doi: 10.1016/j.ypmed.2011.12.024] [Medline: 22244800]
16. West DS, Krukowski RA, Finkelstein EA, Stansbury ML, Ogden DE, Monroe CM, et al. Adding financial incentives to online group-based behavioral weight control: an RCT. Am J Prev Med 2020 Aug;59(2):237-246 [FREE Full text] [doi: 10.1016/j.amepre.2020.03.015] [Medline: 32446752]
17. Kullgren JT, Troxel AB, Loewenstein G, Asch DA, Norton LA, Wesby L, et al. Individual- versus group-based financial incentives for weight loss: a randomized, controlled trial. Ann Intern Med 2013 Apr 02;158(7):505-514 [FREE Full text] [doi: 10.7326/0003-4819-158-7-201304020-00002] [Medline: 23546562]




XSL•FO
RenderX

18. Volpp KG, John LK, Troxel AB, Norton L, Fassbender J, Loewenstein G. Financial incentive-based approaches for weight loss: a randomized trial. JAMA 2008 Dec 10;300(22):2631-2637 [FREE Full text] [doi: 10.1001/jama.2008.804] [Medline: 19066383]
19. Finkelstein EA, Linnan LA, Tate DF, Birken BE. A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees. J Occup Environ Med 2007 Sep;49(9):981-989. [doi: 10.1097/JOM.0b013e31813c6dcb] [Medline: 17848854]
20. Wilson SM, Newins AR, Medenblik AM, Kimbrel NA, Dedert EA, Hicks TA, et al. Contingency management versus psychotherapy for prenatal smoking cessation: a meta-analysis of randomized controlled trials. Womens Health Issues 2018;28(6):514-523 [FREE Full text] [doi: 10.1016/j.whi.2018.05.002] [Medline: 30061033]
21. Higgins ST, Washio Y, Heil SH, Solomon LJ, Gaalema DE, Higgins TM, et al. Financial incentives for smoking cessation among pregnant and newly postpartum women. Prev Med 2012 Nov;55 Suppl:S33-S40 [FREE Full text] [doi: 10.1016/j.ypmed.2011.12.016] [Medline: 22227223]
22. Phillips JK, Skelly JM, Roberts LM, Bernstein IM, Higgins ST. Combined financial incentives and behavioral weight management to enhance adherence with gestational weight gain guidelines: a randomized controlled trial. Am J Obstet Gynecol MFM 2019 Mar;1(1):42-49. [doi: 10.1016/j.ajogmf.2019.02.002] [Medline: 33319756]
23. Paul-Ebhohimhen V, Avenell A. Systematic review of the use of financial incentives in treatments for obesity and overweight. Obes Rev 2008 Jul;9(4):355-367. [doi: 10.1111/j.1467-789X.2007.00409.x] [Medline: 17956546]
24. Volpp KG, Asch DA, Galvin R, Loewenstein G. Redesigning employee health incentives--lessons from behavioral economics. N Engl J Med 2011 Aug 04;365(5):388-390 [FREE Full text] [doi: 10.1056/NEJMp1105966] [Medline: 21812669]
25. Phelan S, Jankovitz K, Hagobian T, Abrams B. Reducing excessive gestational weight gain: lessons from the weight control literature and avenues for future research. Womens Health (Lond) 2011 Nov;7(6):641-661 [FREE Full text] [doi: 10.2217/whe.11.70] [Medline: 22040207]
26. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265 [FREE Full text] [doi: 10.1002/oby.20946] [Medline: 25521523]
27. Lombard C, Harrison C, Kozica S, Zoungas S, Ranasinha S, Teede H. Preventing weight gain in women in rural communities: a cluster randomised controlled trial. PLoS Med 2016 Jan;13(1):e1001941 [FREE Full text] [doi: 10.1371/journal.pmed.1001941] [Medline: 26785406]
28. Vanwormer JJ, French SA, Pereira MA, Welsh EM. The impact of regular self-weighing on weight management: a systematic literature review. Int J Behav Nutr Phys Act 2008 Nov 04;5:54 [FREE Full text] [doi: 10.1186/1479-5868-5-54] [Medline: 18983667]
29. Linde JA, Jeffery RW, French SA, Pronk NP, Boyle RG. Self-weighing in weight gain prevention and weight loss trials. Ann Behav Med 2005 Dec;30(3):210-216. [doi: 10.1207/s15324796abm3003_5] [Medline: 16336072]
30. Shin DW, Yun JM, Shin J, Kwon H, Min HY, Joh H, et al. Enhancing physical activity and reducing obesity through smartcare and financial incentives: a pilot randomized trial. Obesity (Silver Spring) 2017 Feb;25(2):302-310. [doi: 10.1002/oby.21731] [Medline: 28063226]
31. Ries A, Blackman L, Page R, Gizlice Z, Benedict S, Barnes K, et al. Goal setting for health behavior change: evidence from an obesity intervention for rural low-income women. Rural Remote Health 2014;14:2682 [FREE Full text] [Medline: 24785265]
32. Jeffery R, Wing RR, Sherwood NE, Tate DF. Physical activity and weight loss: does prescribing higher physical activity goals improve outcome? Am J Clin Nutr 2003 Oct;78(4):684-689. [doi: 10.1093/ajcn/78.4.684] [Medline: 14522725]
33. Brown MJ, Sinclair M, Liddle D, Hill AJ, Madden E, Stockdale J. A systematic review investigating healthy lifestyle interventions incorporating goal setting strategies for preventing excess gestational weight gain. PLoS One 2012;7(7):e39503 [FREE Full text] [doi: 10.1371/journal.pone.0039503] [Medline: 22792178]
34. Volpp KG, Pauly MV, Loewenstein G, Bangsberg D. P4P4P: an agenda for research on pay-for-performance for patients. Health Aff (Millwood) 2009;28(1):206-214 [FREE Full text] [doi: 10.1377/hlthaff.28.1.206] [Medline: 19124872]
35. O'Donoghue T, Rabin M. The economics of immediate gratification. J Behav Decis Making 2000 Apr;13(2):233-250. [doi: 10.1002/(sici)1099-0771(200004/06)13:2<233::aid-bdm325>3.0.co;2-u]
36. Loewenstein G, Asch DA, Volpp KG. Behavioral economics holds potential to deliver better results for patients, insurers, and employers. Health Aff (Millwood) 2013 Jul;32(7):1244-1250. [doi: 10.1377/hlthaff.2012.1163] [Medline: 23836740]
37. Patel MS, Asch DA, Rosin R, Small DS, Bellamy SL, Heuer J, et al. Framing financial incentives to increase physical activity among overweight and obese adults: a randomized, controlled trial. Ann Intern Med 2016 Mar 15;164(6):385-394 [FREE Full text] [doi: 10.7326/M15-1635] [Medline: 26881417]
38. Leahey TM, Subak LL, Fava J, Schembri M, Thomas G, Xu X, et al. Benefits of adding small financial incentives or optional group meetings to a web-based statewide obesity initiative. Obesity (Silver Spring) 2015 Jan;23(1):70-76 [FREE Full text] [doi: 10.1002/oby.20937] [Medline: 25384463]
39. Kanfer FH, Goldstein AP. Helping People Change: A Textbook of Methods. Oxford, United Kingdom: Pergamon Press; 1986.



XSL•FO
RenderX

40. Gardner B, Wardle J, Poston L, Croker H. Changing diet and physical activity to reduce gestational weight gain: a meta-analysis. Obes Rev 2011 Jul;12(7):e602-e620. [doi: 10.1111/j.1467-789X.2011.00884.x] [Medline: 21521451]
41. Asch DA, Volpp KG. On the way to health. LDI Issue Brief 2012;17(9):1-4. [Medline: 22934330]
42. Sen AP, Sewell TB, Riley EB, Stearman B, Bellamy SL, Hu MF, et al. Financial incentives for home-based health monitoring: a randomized controlled trial. J Gen Intern Med 2014 May 13;29(5):770-777 [FREE Full text] [doi: 10.1007/s11606-014-2778-0] [Medline: 24522623]
43. Thaler RH, Tversky A, Kahneman D, Schwartz A. The effect of myopia and loss aversion on risk taking: an experimental test. Quarterly J Econ 1997 May 01;112(2):647-661. [doi: 10.1162/003355397555226]
44. Committee on Obstetric Practice. Committee Opinion No. 650 Physical Activity and Exercise During Pregnancy and the Postpartum Period. Obstet Gynecol 2015 Dec;126(6):e135-e142. [doi: 10.1097/aog.0000000000001214]
45. Cadmus-Bertram L, Marcus BH, Patterson RE, Parker BA, Morey BL. Use of the Fitbit to measure adherence to a physical activity intervention among overweight or obese, postmenopausal women: self-monitoring trajectory during 16 weeks. JMIR Mhealth Uhealth 2015 Nov 19;3(4):e96 [FREE Full text] [doi: 10.2196/mhealth.4229] [Medline: 26586418]
46. Oken E, Taveras EM, Kleinman KP, Rich-Edwards JW, Gillman MW. Gestational weight gain and child adiposity at age 3 years. Am J Obstet Gynecol 2007 Apr;196(4):322.e1-322.e8 [FREE Full text] [doi: 10.1016/j.ajog.2006.11.027] [Medline: 17403405]
47. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
48. Collins L. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham: Springer; 2018.
49. Harrison CL, Teede HJ, Lombard CB. How effective is self-weighing in the setting of a lifestyle intervention to reduce gestational weight gain and postpartum weight retention? Aust N Z J Obstet Gynaecol 2014 Aug;54(4):382-385. [doi: 10.1111/ajo.12207] [Medline: 24738837]
50. Finkelstein EA, Tham K, Haaland BA, Sahasranaman A. Applying economic incentives to increase effectiveness of an outpatient weight loss program (TRIO) - A randomized controlled trial. Soc Sci Med 2017 Jul;185:63-70. [doi: 10.1016/j.socscimed.2017.05.030] [Medline: 28554160]
51. Giles EL, Becker F, Ternent L, Sniehotta FF, McColl E, Adams J. Acceptability of financial incentives for health behaviours: a discrete choice experiment. PLoS One 2016;11(6):e0157403 [FREE Full text] [doi: 10.1371/journal.pone.0157403] [Medline: 27314953]
52. McGill B, O'Hara BJ, Grunseit AC, Bauman A, Osborne D, Lawler L, et al. Acceptability of financial incentives for maintenance of weight loss in mid-older adults: a mixed methods study. BMC Public Health 2018 Feb 13;18(1):244 [FREE Full text] [doi: 10.1186/s12889-018-5136-z] [Medline: 29439689]
53. Pereira MA, Rifas-Shiman SL, Kleinman KP, Rich-Edwards JW, Peterson KE, Gillman MW. Predictors of change in physical activity during and after pregnancy: project viva. Am J Prev Med 2007 Apr;32(4):312-319 [FREE Full text] [doi: 10.1016/j.amepre.2006.12.017] [Medline: 17383562]
54. Choi J, Fukuoka Y, Lee JH. The effects of physical activity and physical activity plus diet interventions on body weight in overweight or obese women who are pregnant or in postpartum: a systematic review and meta-analysis of randomized controlled trials. Prev Med 2013 Jun;56(6):351-364 [FREE Full text] [doi: 10.1016/j.ypmed.2013.02.021] [Medline: 23480971]
55. Krukowski RA, West D, DiCarlo M, Shankar K, Cleves MA, Tedford E, et al. A behavioral intervention to reduce excessive gestational weight gain. Matern Child Health J 2017 Mar;21(3):485-491. [doi: 10.1007/s10995-016-2127-5] [Medline: 27449652]
56. Chen MS, Lara PN, Dang JH, Paterniti DA, Kelly K. Twenty years post-NIH Revitalization Act: enhancing minority participation in clinical trials (EMPaCT): laying the groundwork for improving minority clinical trial accrual: renewing the case for enhancing minority participation in cancer clinical trials. Cancer 2014 Apr 01;120 Suppl 7:1091-1096 [FREE Full text] [doi: 10.1002/cncr.28575] [Medline: 24643646]
57. Haughton CF, Silfee VJ, Wang ML, Lopez-Cepero AC, Estabrook DP, Frisard C, et al. Racial/ethnic representation in lifestyle weight loss intervention studies in the United States: a systematic review. Prev Med Rep 2018 Mar;9:131-137 [FREE Full text] [doi: 10.1016/j.pmedr.2018.01.012] [Medline: 29616185]
58. Renjilian DA, Perri MG, Nezu AM, McKelvey WF, Shermer RL, Anton SD. Individual versus group therapy for obesity: effects of matching participants to their treatment preferences. J Consult Clin Psychol 2001 Aug;69(4):717-721. [Medline: 11550739]
59. Hill B, Skouteris H, Fuller-Tyszkiewicz M. Interventions designed to limit gestational weight gain: a systematic review of theory and meta-analysis of intervention components. Obes Rev 2013 Jun;14(6):435-450. [doi: 10.1111/obr.12022] [Medline: 23534901]
60. Phelan S, Wing RR, Brannen A, McHugh A, Hagobian TA, Schaffner A, et al. Randomized controlled clinical trial of behavioral lifestyle intervention with partial meal replacement to reduce excessive gestational weight gain. Am J Clin Nutr 2018 Feb 01;107(2):183-194 [FREE Full text] [doi: 10.1093/ajcn/nqx043] [Medline: 29529157]




XSL•FO
RenderX

## Abbreviations

**GWG:** gestational weight gain
**IOM:** Institute of Medicine

*Edited by G Eysenbach; submitted 26.05.21; peer-reviewed by N Meloncelli, S Hertling, HL Tam; comments to author 02.07.21; revised version received 29.07.21; accepted 17.11.21; published 24.12.21.*

*Please cite as:*
*Krukowski R, Johnson B, Kim H, Sen S, Homsi R*
*A Pragmatic Intervention Using Financial Incentives for Pregnancy Weight Management: Feasibility Randomized Controlled Trial*
*JMIR Form Res 2021;5(12):e30578*
*URL: https://formative.jmir.org/2021/12/e30578*
*doi:10.2196/30578*
*PMID:34951594*

©Rebecca Krukowski, Brandi Johnson, Hyeonju Kim, Saunak Sen, Riad Homsi. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# A Smartphone-Based Self-management Intervention for Individuals With Bipolar Disorder (LiveWell): Protocol Development for an Expert System to Provide Adaptive User Feedback

Evan H Goulding[1], MD, PhD; Cynthia A Dopke[1], PhD; Tania Michaels[2], MD; Clair R Martin[1], BA; Monika A Khiani[3], MD; Christopher Garborg[1], BSc; Chris Karr[4], MA; Mark Begale[5], BA

[1]Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Deparment of Pediatrics, Loma Linda Children's Hospital, Loma Linda, CA, United States

[3]Rogers Behavioral Health, Milwaukee, WI, United States

[4]Audacious Software, Chicago, IL, United States

[5]Vibrent Health, Fairfax, VA, United States

**Corresponding Author:**
Evan H Goulding, MD, PhD
Department of Psychiatry and Behavioral Sciences
Feinberg School of Medicine
Northwestern University
Suite 7-102
303 East Chicago Ave
Chicago, IL, 60611
United States
Phone: 1 312 503 1189
Email: e-goulding@fsm.northwestern.edu

## Abstract

**Background:** Bipolar disorder is a severe mental illness that results in significant morbidity and mortality. While pharmacotherapy is the primary treatment, adjunctive psychotherapy can improve outcomes. However, access to therapy is limited. Smartphones and other technologies can increase access to therapeutic strategies that enhance self-management while simultaneously augmenting care by providing adaptive delivery of content to users as well as alerts to providers to facilitate clinical care communication. Unfortunately, while adaptive interventions are being developed and tested to improve care, information describing the components of adaptive interventions is often not published in sufficient detail to facilitate replication and improvement of these interventions.

**Objective:** To contribute to and support the improvement and dissemination of technology-based mental health interventions, we provide a detailed description of the expert system for adaptively delivering content and facilitating clinical care communication for LiveWell, a smartphone-based self-management intervention for individuals with bipolar disorder.

**Methods:** Information from empirically supported psychotherapies for bipolar disorder, health psychology behavior change theories, and chronic disease self-management models was combined with user-centered design data and psychiatrist feedback to guide the development of the expert system.

**Results:** Decision points determining the timing of intervention option adaptation were selected to occur daily and weekly based on self-report data for medication adherence, sleep duration, routine, and wellness levels. These data were selected for use as the tailoring variables determining which intervention options to deliver when and to whom. Decision rules linking delivery of options and tailoring variable thresholds were developed based on existing literature regarding bipolar disorder clinical status and psychiatrist feedback. To address the need for treatment adaptation with varying clinical statuses, decision rules for a clinical status state machine were developed using self-reported wellness rating data. Clinical status from this state machine was incorporated into hierarchal decision tables that select content for delivery to users and alerts to providers. The majority of the adaptive content addresses sleep duration, medication adherence, managing signs and symptoms, building and utilizing support, and keeping a regular routine, as well as determinants underlying engagement in these target behaviors as follows: attitudes and perceptions, knowledge, support, evaluation, and planning. However, when problems with early warning signs, symptoms, and transitions to more acute clinical states are detected, the decision rules shift the adaptive content to focus on managing signs and symptoms, and engaging with psychiatric providers.



XSL•FO
RenderX

**Conclusions:** Adaptive mental health technologies have the potential to enhance the self-management of mental health disorders. The need for individuals with bipolar disorder to engage in the management of multiple target behaviors and to address changes in clinical status highlights the importance of detailed reporting of adaptive intervention components to allow replication and improvement of adaptive mental health technologies for complex mental health problems.

***(JMIR Form Res 2021;5(12):e32932)*** doi:10.2196/32932

**KEYWORDS**

adaptive; personalized; self-management; smartphone; behavioral intervention technology; mHealth; bipolar disorder; depression; mania

## Introduction

Bipolar disorder is a severe mental illness characterized by episodes of mania, hypomania, depression, and mixed states [1,2]. Pharmacotherapy is the primary treatment for bipolar disorder, but even when pharmacological treatment is initially effective, high rates of episode recurrence, interepisode symptoms, and psychosocial impairment persist [3-7]. Combining psychotherapy with pharmacotherapy decreases episode recurrence and symptom burden while also improving quality of life [8-16], and treatment guidelines recommend providing access to psychotherapy for individuals with bipolar disorder [17-19]. Despite these recommendations and the demonstrated effectiveness of adjunctive psychotherapy, multiple barriers limit access to therapy, and only half of individuals with bipolar disorder receive any psychotherapy [20-24]. Enhancing access to tools and content derived from empirically supported psychotherapies thus has the potential to improve treatment for individuals with bipolar disorder.

Because smartphones are widely used and accepted for mental health assistance [25-29], smartphone-based mental health technologies (MHTs) provide a promising means for increasing access to tools and content derived from psychotherapy. In addition, individuals with bipolar disorder in sustained remission report using self-management strategies that overlap significantly with those delivered by empirically supported psychotherapies, and many people with bipolar disorder are interested in utilizing self-management strategies [30-33]. This suggests that MHTs delivering strategies derived from empirically supported psychotherapies may meet user needs and support engagement with effective self-management strategies [34-37].

Because psychotherapy can be considered as a dynamic process that entails ongoing assessment and re-evaluation of an individual's evolving needs and health status [38-41], MHTs also provide novel opportunities for improving the use of self-management strategies [42-45]. Relative to face-to-face treatment, MHTs can deliver real-time assessments in the context of individuals' day-to-day life unrestricted by clinical appointment frequency. This assessment information can then be used to adapt the frequency, mode, and content of user support. Tailoring intervention content to individuals has been shown to improve intervention outcomes [46-48], and adaptive interventions that vary intervention options to meet the changing needs of individuals are being investigated [49-51]. These adaptive interventions use fixed decision rules to link assessment of tailoring variables with the delivery of support to provide a flexible intervention design while maintaining replicability [49-51]. Unfortunately, most recently published adaptive intervention protocols have not provided detailed reporting of intervention components required to enable replication [52].

To address the need for increased access to and enhancement of empirically supported tools and content for individuals with bipolar disorder, a novel smartphone-based self-management intervention (LiveWell) has been developed (NCT02405117) and tested in a single-blind randomized controlled trial (NCT03088462). Because adequate description of interventions is essential to facilitate ongoing efforts to improve and disseminate empirically supported treatments [53-56], the adaptive components (decision points, tailoring variables, decision rules, mode, and content of delivery options) for the LiveWell intervention are described here. The overall intervention framework and design for LiveWell, including the delivery and timing of the fixed content and the evaluation methodology, are described in detail elsewhere [57].

## Methods

To assist in developing the expert system providing LiveWell's adaptive content, an electronic survey (Multimedia Appendix 1) approved by the Northwestern University Institutional Review Board was sent to psychiatrists (N=161) at university-affiliated and private outpatient mental health practices. The survey aimed to obtain psychiatrist feedback on when the smartphone app should prompt users to contact their psychiatrists and when the app should additionally send an email alert to the psychiatrist. In addition, information was requested concerning what type of data would be most useful in a web accessible report for psychiatrists and preferences for being contacted by users. To assist in developing the decision rules, modes and cumulative percent responses were determined from the completed surveys (N=42, 26% response rate). Percent responses were also calculated for preferences and opinions concerning web accessible reports and contact preferences (Multimedia Appendix 2).

To quantify the content of the adaptive delivery options, the behavior change framework that guided the creation of the content and tools for LiveWell was used to code every page of the smartphone app [57]. This framework proposes that (1) engaging in target behaviors improves clinical and recovery outcomes, (2) behavioral determinants govern enactment of target behaviors, and (3) exposure to behavior change technique content and tool use alters behavioral determinants. Determinants and their corresponding techniques are grouped



XSL•FO
RenderX

into the following 4 domains: motivational determinants involved in developing an intention to engage in a behavior, volitional determinants involved in enacting the behavior, and environmental determinants and capabilities that impact motivational and volitional processes. This framework provides a means to label app content in terms of outcomes, targets, and determinants addressed by the behavior change techniques delivered on each app page. To quantify variation in the adaptive content delivered to users based on their current assessment data and the decision rules (Multimedia Appendix 3), app content was exported to Excel spreadsheets (Microsoft Corp) for labeling, and the labeled content was processed using custom code written and run using MATLAB (MathWorks) [57].

## Results

### Intervention Overview

The LiveWell intervention aims to decrease episode relapse, reduce symptom burden, and improve quality of life by assisting individuals with managing target behaviors proposed to underlie the impact of existing therapies [57-60]. LiveWell therefore engages users to support managing the signs and symptoms of relapse, taking psychiatric medications as prescribed, obtaining adequate sleep duration, and maintaining regular routines. LiveWell also addresses strengthening social support, managing stressors, and engaging in healthy habits regarding diet, exercise, and substance use.

The intervention consists of technological and human support components including a smartphone app, a secure server and website, and a coach [57-60]. The smartphone app consists of the following 5 primary components: *foundations*, *toolbox*, *wellness plan*, *daily check-in*, and *daily review* [57-60]. The app provides information on bipolar disorder self-management (*foundations*) along with self-assessment surveys and skills content (*toolbox*). The *foundations* and *toolbox* components support developing a personalized *wellness plan* for managing signs and symptoms and maintaining a healthy lifestyle. The core of the intervention is a *daily check-in*, where users monitor medication adherence, sleep duration, routine, and wellness levels. Based on their *daily check-in* data, *daily review* provides adaptive content and directs users to relevant app content in the *foundations*, *toolbox*, and *wellness plan*. A coach provides human support to facilitate app use adherence, self-management strategy use, and communication with mental health providers [58]. The secure server and website provide data summaries and alerts to providers and coaches to facilitate and support clinical care communication (Multimedia Appendix 4).

The LiveWell intervention thus has both fixed and adaptive components. In terms of fixed components, users are asked to attend an initial face-to-face meeting with a coach, complete 6 scheduled coaching calls (weeks 1-4, 6, and 16), read 2 *foundations* lessons per week (weeks 1-4), and use the *daily check-in* each day (weeks 1-16). At an initial face-to-face meeting, a coach helps users identify personalized wellness anchors to assist them in using the *daily check-in's* wellness rating scale [58,59]. During the scheduled coach calls, users review progress toward their goals and receive guidance on app use (weeks 1-4), develop (week 4) and review (week 6) their personalized wellness plan, and then review their progress and commitment to ongoing use of helpful strategies (week 16). The week 16 scheduled call ends the active phase of the intervention, but users continue to have access to the app and ad hoc coaching support for 48 weeks. While the target behavioral goals and wellness plans are personalized, this personalization is determined by the user in conjunction with the coach without following a predefined algorithm and is thus operator driven and not algorithmically driven [57-60].

The adaptive components of the intervention consist of a rule-based expert system primarily embedded in the *daily check-in* and *daily review* components of the intervention. The *daily check-in* provides the primary user interface for populating the database containing facts about the user's current status, but data are also collected via a *weekly check-in* consisting of the Patient Health Questionnaire-8 (PHQ8) and Altman Self-Rating Mania (ASRM) scales [61,62]. The knowledge base for the expert system contains the adaptive smartphone app content. It utilizes a clinical status state machine and hierarchical decision tables (if-then/elseif-then) to govern content delivery via the *daily review*. The *daily review* comprises the explanatory system and also serves as the user interface providing adaptive content delivery. The inference engine combines the current data from the user database with the knowledge base to determine the adaptive content to deliver in the moment. In addition to varying the content provided via the *daily review*, the expert system also controls the delivery of smartphone app pop-up messages requesting users to reach out to their clinical care providers by phone. Additionally, the expert system controls the delivery of emails to providers and coaches to stimulate clinical care communication via phone calls, and flags data in daily reports generated for coaches to promote coach clinical reach out and app use adherence phone calls [57-59].

### Decision Points

Decision points are the times when the frequency, mode, or content of the intervention (ie, intervention options) is adapted [49,51,52]. For LiveWell, the decision points occur daily upon user completion of the *daily check-in*, as well as weekly upon completion of the *weekly check-in*. During a pilot study of the intervention, users were initially given the option to complete the *daily check-in* multiple times a day. However, allowing multiple check-ins each day did not appear to elicit user reflection about their wellness status. Instead, multiple check-ins a day seemed to be capturing momentary reactions to daily hassles and uplifts [63]. To encourage users to engage in reflective monitoring rather than in the moment rating, the *daily check-in* was restricted to allow only 1 check-in per day [59]. The app prompts users to check-in each day by delivering an alert in the smartphone's notification panel. Users can select the time they first receive a daily notification in the app settings. They receive their first notification to check-in at their designated time of the day and then every 2 hours until they check-in or until they receive 3 alerts. The *dailyreview* is only accessible each day after completion of the *daily check-in*. As a result, no adaptive content is delivered via the *daily review* on days when users do not complete the *daily check-in*. Similarly, no adaptive content is delivered if users do not complete the *weekly check-in*.



XSL•FO
RenderX

### Tailoring Variables

Tailoring variables are the values containing information about the user, such as clinical status or health behaviors, which are used to determine the delivery of intervention options [49,51,52]. For LiveWell, the primary tailoring variables are the target behaviors and wellness levels monitored by the *daily check-in* and include users self-reported psychiatric medication adherence (all, some, and none), sleep duration (hours), routine (bedtime/risetime), and wellness rating (0 balanced, −1/+1 daily hassles/uplifts, −2/+2 prodromal/residual symptoms, −3/+3 episode, and −4/+4 crisis). These variables were selected for daily monitoring because they are consistently addressed in the core content of adjunctive psychotherapy interventions [3,16,24,31,64-66] and are readily amenable to goal setting and self-monitoring. LiveWell also contains content addressing building and using supports, managing stressors, and maintaining healthy habits regarding diet, exercise, and substance use [57-60]. However, to simplify the *daily check-in* and make it quick and easy to complete, these additional targets were not included for daily monitoring and use in adapting delivery of intervention options. Because of the importance of clinical status in bipolar disorder, 2 additional tailoring variables were also utilized based on data obtained from the *weekly check-in* (PHQ8 and ASRM scores). In addition, adherence to completing the *daily* and *weekly check-ins* was also used as a tailoring variable.

### Decision Rules

Decision rules link delivery of intervention options and tailoring variables systematically. They include the values of tailoring variables (states, thresholds, and ranges) that determine which intervention option to deliver when and for whom [49,51,52]. LiveWell utilizes algorithm-driven adaptation based on fixed rules to standardize intervention option delivery [52]. In developing the decision rules for LiveWell, it was necessary to consider how variation in medication adherence, sleep duration, routine, and wellness ratings should result in the adaptation of intervention options. In addition, it was also necessary to consider how to rank the importance of variation in each variable. Because pharmacotherapy is the primary treatment for bipolar disorder [3-7], medication adherence was ranked above sleep duration in prioritizing feedback. Sleep duration was then ranked above routine because sleep duration variation is frequently identified as an early warning sign of impending episode relapse, and assisting people with identifying and making plans for managing early warning signs of relapse is an important target of most adjunctive bipolar disorder psychotherapies [67-69]. In fact, many face-to-face bipolar disorder psychotherapy studies have restricted participation to individuals in asymptomatic recovery, and the primary goal of these interventions has been relapse prevention [8-13,16]. However, an important goal of LiveWell is to increase access to self-management techniques derived from empirically supported therapies, so individuals were enrolled in the intervention as long as they were not in a current episode [57]. Because individuals with varying clinical statuses were included initially and people may enter an episode while engaged with LiveWell, clinical status was ranked as the highest priority tailoring variable. This allows the need for substantial adaptation of the treatment approach for individuals with bipolar disorder in various clinical states to be appropriately addressed. As a starting point for developing decision rules to manage clinical status variation, existing criteria for bipolar disorder episodes (mania, hypomania, depression, and mixed) and for nonepisode clinical statuses (asymptomatic and symptomatic recovery, prodromal, continued symptomatic, and recovering) were examined (Table 1 and Table 2). The Diagnostic and Statistical Manual of Mental Disorders fourth edition (DSM4) criteria were used because prior face-to-face psychotherapy trials for bipolar disorder primarily used DSM4 episode criteria [8-16]. They were used for LiveWell's outcome assessments [57]. Criteria for nonepisode clinical statuses were derived from the clinical monitoring form (CMF) used in a large bipolar disorder intervention study (STEP-BD) where nonepisode clinical statuses were clearly defined [70,71]. These combined criteria were used to define a clinical status state machine where transitions between the 9 possible clinical statuses for bipolar disorder are defined by 10 decision rules derived from DSM4 and the CMF (Figure 1).

The DSM4/CMF state machine highlights the relatively narrow focus of most adjunctive psychotherapies for bipolar disorder (Figure 1). These therapies primarily assist people with remaining in asymptomatic recovery by maintaining medication adherence, adequate sleep duration, and a regular routine, as well as attending to healthy habits and stressors [3,8-13,16,24,31,64,65]. In addition, these therapies assist individuals with recognizing and managing early warning signs and symptoms to avoid transitioning from asymptomatic recovery to a prodromal state and to support transitioning from a prodromal state to asymptomatic recovery rather than entering an episode (Figure 1). This state machine also highlights how entry into an episode can lead to prolonged periods of problems with symptoms; individuals can cycle between episode types, and into and out of continued symptomatic and recovering states during which they are at increased risk for episode relapse [7,72-74]. While this state machine clearly delineates different clinical states and transitions, assessment is complex because of the need to determine the number of symptoms as well as their type and intensity over widely varying durations (4-56 days), and the additional need to assess impairment level and the presence of psychotic symptoms or hospitalization. The DSM4/CMF data necessary to determine the clinical status states for the DSM4/CMF state machine are obtained every 8 weeks during the intervention via telephone assessments delivered by trained assessors blinded to the study arm [57]. However, because the severity of each symptom and the level of impairment due to both symptoms of mania and depression must be assessed, it was decided that this level of assessment via the app would unduly burden the user.



XSL•FO
RenderX

**Table 1.** Diagnostic and Statistical Manual of Mental Disorders fourth edition (DSM4)/clinical monitoring form (CMF): episode criteria.

| Clinical status (to) | Entry criteria[a] | Number of moderate symptoms of | Impairment | Consecutive days | DN[b] | PSR[c] |
|---|---|---|---|---|---|---|
| Depression | Yes | Depression ≥5 | ≥Moderate | ≥10/14 | 7 | 5-6 |
| Mixed | Yes[d] | Mania and depression[d] | ≥Moderate | ≥7 | 6 | 5-6 |
| Mania | Yes | Mania ≥3 if elevated; mania ≥4 if only irritable | ≥Moderate, hospitalized, or psychosis | ≥7 or hospitalized | 5 | 5-6 |
| Hypomania | Yes | Mania ≥3 if elevated; mania ≥4 if only irritable | <Moderate, not hospitalized, no psychosis | ≥4 | 4 | 3 |

[a]Entry criteria: meeting consecutive day criteria; for mania/hypomania, moderate severity elevated, expansive, or irritable mood; and for depression, moderate severity depressed mood or loss of interest/pleasure. Moderate symptom and impairment criteria based on the clinical monitoring form.

[b]DN: decision rule number.

[c]PSR: psychiatric status rating; 1=no symptoms, 1.5=mild symptoms, 2=residual or prodromal symptoms, 3=moderate symptoms, 4=marked symptoms, 5=episode, and 6=severe episode.

[d]Criteria for mania with concurrent depressive symptoms for 5/7 consecutive days.

**Table 2.** Clinical monitoring form (CMF) clinical status decision rules when not in an episode.

| Clinical status (to) | Recovery | Symptom count[a]/impairment | Consecutive days | DN[b] | PSR[c] |
|---|---|---|---|---|---|
| Continued symptomatic | No | Symptom count >2 or ≥moderate impairment | ≥7[d] | 8 | 3-4 |
| Prodromal | Yes | Symptom count >2 or new[e] or ≥moderate impairment | ≥7[d] | 1 | 2-4 |
| Recovering | No | Symptom count ≤2 and <moderate impairment | ≥7[d] and ≤56 | 9 | 1-2 |
| Recovery | Yes | Symptom count ≤2 and <moderate impairment | >56 | 10 | 1-2 |
| Symptomatic recovery | Yes | Symptom count >0 but ≤2 and <moderate impairment | ≥7[f] | 3 | 2 |
| Asymptomatic recovery | Yes | Symptom count=0 and <moderate impairment | ≥7[f] | 2 | <2 |

[a]Symptom count: sum of symptom severity. If |severity| ≥1, round up; otherwise, 0. Clinical monitoring form (CMF) symptom severity scale: none=0, mild=0.5, moderate=1, marked=1.5, and severe=2.

[b]DN: decision rule number.

[c]PSR: psychiatric status rating; 1=no symptoms, 1.5=mild symptoms, 2=residual or prodromal symptoms, 3=moderate symptoms, 4=marked symptoms, 5=episode, and 6=severe episode.

[d]The 7 consecutive day window was selected to align with the 7-day evaluation window used for the LIFE-CMF assessments.

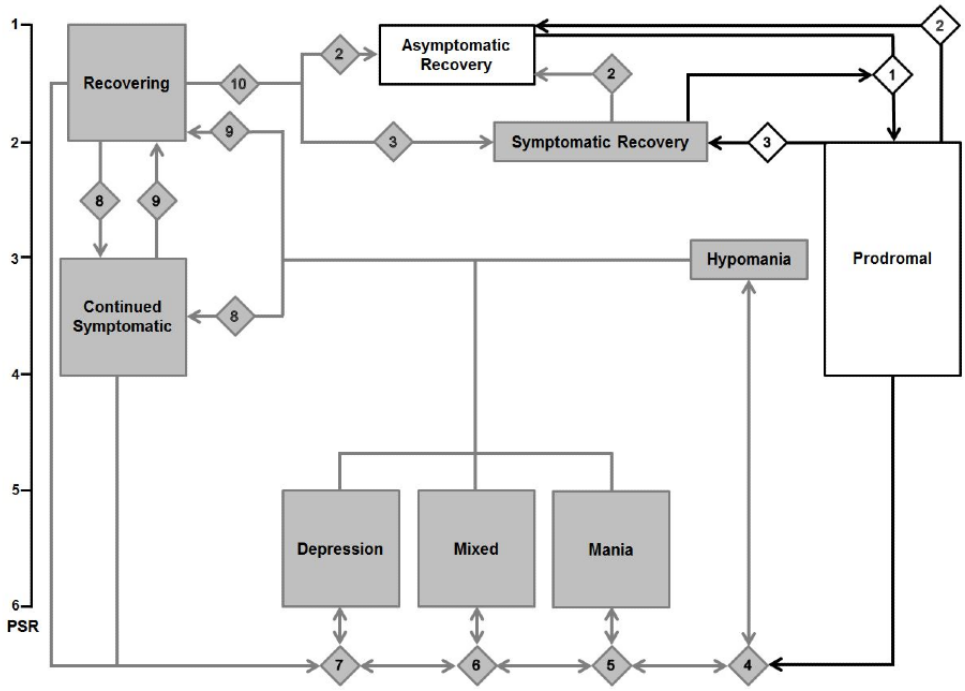
[e]Two new moderate, marked, or severe symptoms developed while in recovery.

[f]From prodromal or symptomatic recovery only and not from recovering.



XSL•FO
RenderX

**Figure 1.** Diagnostic and Statistical Manual of Mental Disorders fourth edition/clinical monitoring form clinical status state machine. Diamonds show decision rule numbers (see Table 1 and Table 2 for rule details). White rectangles, diamonds, and black lines indicate primary clinical status states and transitions addressed by most adjunctive psychotherapies for bipolar disorder and the main focus of the LiveWell intervention. Psychiatric status rating (PSR) is as follows: 1, no symptoms; 1.5, mild symptoms; 2, residual or prodromal symptoms; 3, moderate symptoms; 4, marked symptoms; 5, episode; 6, severe episode.




Thus, the clinical status state machine for the LiveWell expert system was simplified owing to the complexity of the data required for the DSM4/CMF state machine and the corresponding user assessment burden (Table 3 and Figure 2). Instead of having users rate individual symptom severity, the *daily check-in* uses a personalized 9-point wellness rating scale to allow users to rapidly assess their symptom burden using the *daily check-in* [58,59]. The clinical status assessment window was also limited to users' last 7 *daily check-ins* to reflect user self-report data covering approximately the prior week depending on *daily check-in* adherence (Table 3). Consistent with the focus of LiveWell on episode relapse prevention, this time window aims to allow relatively rapid adjustment of intervention option delivery with clinical status variation. In addition to limiting the time window, the LiveWell state machine was also simplified by considering only the following 4 clinical states: well, prodromal, unwell, and recovering (Table 3 and Figure 2). Reducing the number of clinical states allows the *daily check-in* wellness ratings to be readily used to identify clinical status state transitions using the same time window for symptoms of depression and mania, thereby reducing the number of decision rules needed and providing more rapid feedback to users when they identify changes in their wellness ratings. In addition, the focus of the LiveWell intervention is on mood episode relapse prevention. The intervention thus aims to assist individuals who have recovered from a mood episode on transitioning from well to prodromal and from prodromal to unwell. If an individual enters a mood episode, the intervention aims to support the individual in working with their providers for treatment. For individuals in the recovering state, the LiveWell intervention adapts techniques developed primarily for mood episode relapse prevention [3,8-13,16,24,31,64,65], with the aim of assisting individuals in transitioning from recovering to well and avoiding transitioning from recovering to unwell. The use of only 4 clinical states for the LiveWell state machine is thus consistent with the focus of the intervention and existing studies from which the intervention content is derived.

**Table 3.** LiveWell clinical status decision rules.

| Clinical status (from) | Clinical status (to) | Wellness rating | | Decision number |
|---|---|---|---|---|
| | | Criteria | Count[a] | |
| Well | Prodromal | \|WR\|[b] ≥2 | ≥4 | 1 |
| Prodromal or recovering | Well | \|WR\| ≤1 | ≥5 | 2 |
| Prodromal or recovering | Unwell | \|WR\| ≥3 | ≥5 | 3 |
| Unwell | Recovering | \|WR\| ≤2 | ≥5 | 4 |

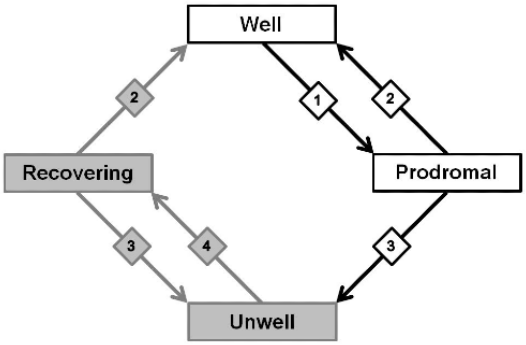
[a]Count of the last 7 daily check-ins meeting the wellness rating criteria.

[b]|WR|: absolute value of daily check-in wellness ratings.



XSL•FO
RenderX

**Figure 2.** LiveWell clinical status state machine. Diamonds show decision rule numbers (see Table 3 for rule details). White rectangles are the primary states addressed by most adjunctive psychotherapies for bipolar disorder.




To aid in defining the thresholds for varying clinical states and to address thresholds for medication adherence and sleep duration, an online electronic survey was sent to practicing psychiatrists to obtain their input on when intervention options should be varied (Multimedia Appendix 1). This questionnaire emphasized identifying the number of days after which psychiatrists felt individuals with bipolar disorder should shift from relying mostly on individual self-management skills and instead engage in self-management via communication and recruiting support from mental health providers. The survey thus asked when the app should deliver a prompt to users requesting that they call their psychiatrist and when the app, via the server, should send an email alert directly notifying the psychiatrist of potential issues (Multimedia Appendix 1). Most psychiatrists indicated that the app should instruct users to call them after 2 to 3 days of experiencing multiple symptoms of mania and after 3 to 4 days of experiencing multiple symptoms of depression or 1 to 2 early warning signs of mania or depression (Multimedia Appendix 2 and Multimedia Appendix 5). In terms of medication adherence, most psychiatrists indicated that the app should instruct users to call them if they reported only taking their psychiatric medications for 3 out of 7 days (43% adherence). For sleep duration, most psychiatrists indicated that the app should instruct users to call them if they reported sleeping less than 4 hours a day for 2 to 3 days and if sleeping less than usual by 3 to 4 hours or more than usual by 4 hours over 7 days. The results were less clear with regard to when psychiatrists would want to receive an email alert (Multimedia Appendix 5), which may be related to most psychiatrists (80%) preferring that their patients contact them via phone rather than via email (Multimedia Appendix 2).

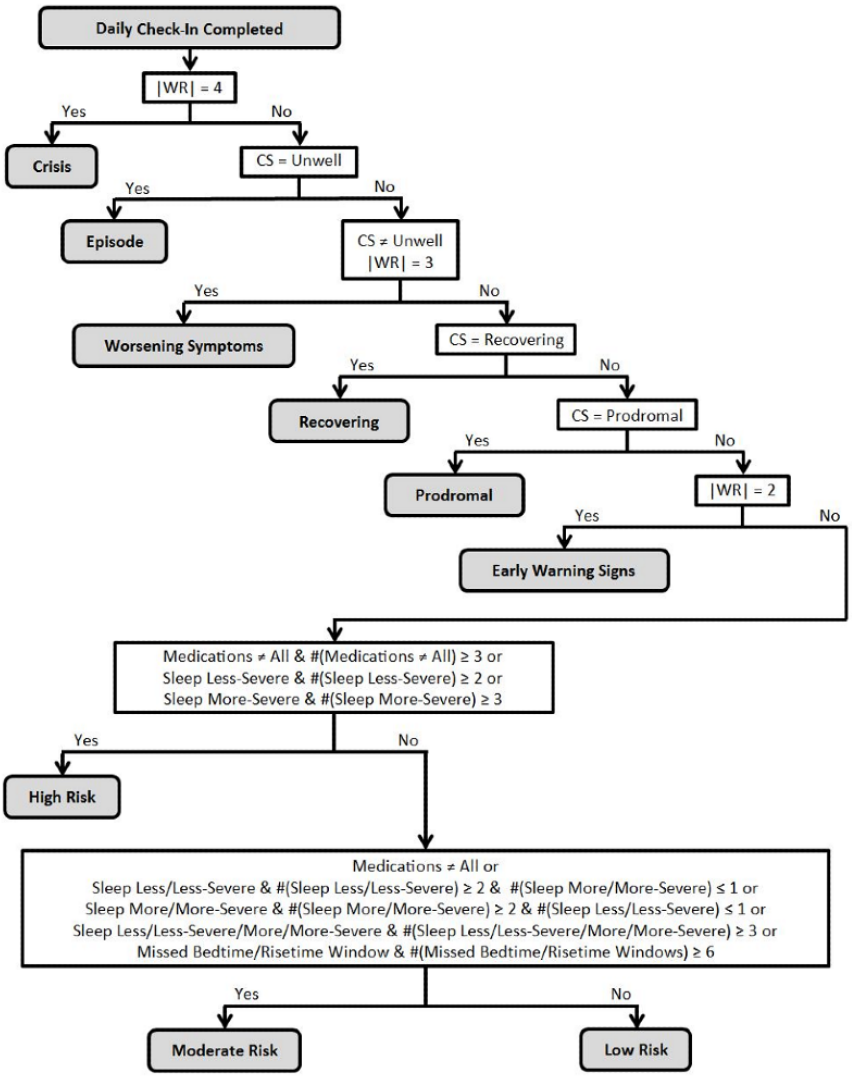
After each completion of the *daily check-in*, the LiveWell inference engine assesses the user's current clinical status and last 7 *daily check-in* wellness ratings and applies the knowledge base rules to update the user's clinical status (Table 3 and Figure 2). The user's initial clinical status is set by the coach using baseline study assessment data and is then updated after each check in. For example, if a user's clinical status is well and the absolute value of their wellness ratings are 2 or greater for 4 or more of the last 7 check-ins, the user's clinical status is updated to prodromal (decision number 1 in Table 3, Figure 2). The coach assists users in understanding and identifying personalized anchors indicating that a wellness rating of +2 or −2 corresponds to the presence of early warning signs or symptoms when doing well or some residual symptoms when improving from being in an episode [58]. The threshold of 4 check-ins corresponds to the DSM4 criteria for hypomania (Table 1) and to psychiatrist feedback from the survey indicating psychiatrists would like their patients to contact them when experiencing manic or depressive symptoms or early warning signs (Multimedia Appendix 5). For the transition from prodromal or recovering to unwell, a count of 5 or more for the last 7 check-ins with an absolute wellness rating value of 3 or greater was used (decision number 3 in Table 3, Figure 2). A wellness rating of 3 is defined for users as corresponding to multiple symptoms day to day, with some impairment likely indicating a mood episode. The threshold of 5 counts out of 7 corresponds to the CMF threshold for a depressive episode using DSM4 criteria (Table 1), which captures the DSM4/CMF criteria of depressive symptoms being present most days for 2 weeks (ie, 10/14 days). However, for the LiveWell clinical status state machine, the count has been reduced by half to reflect the approximately 1 week window of the state machine (ie, 5/7 days).

After the state machine determines clinical status, a hierarchical decision table is used to decide which of the 26 content categories should be delivered by the *daily review* (Figure 3, Multimedia Appendix 3, and Multimedia Appendix 6) [57]. Because clinical status and wellness ratings are prioritized, the *daily review* decision table starts by examining these variables and proceeds using an if-then/elseif-then decision tree. Thus, if a user enters a wellness rating of +4 or −4, the crisis content categories severe up or down are selected. After these 2 categories, the next 15 content categories are selected by examining the clinical status and the current wellness rating (Multimedia Appendix 3 and Multimedia Appendix 6). These content categories address continuing or improving in the clinical states unwell, prodromal, and recovering, as well as worsening symptoms (wellness ratings +3 or −3) while not in the unwell state (well, prodromal, and recovering), and the presence of early warning signs (wellness ratings +2 or −2) while in the well state (Figure 3). The decision rules for the remaining 9 categories examine user self-report data regarding medication adherence, sleep duration, and routine using data from the 4 last check-ins (Figure 3). This approximately 4-day window for deciding when to deliver content addressing these targets was selected based on the psychiatrist survey feedback (Multimedia Appendix 5).



XSL•FO
RenderX

**Figure 3.** Overview of the daily review decision rules. Only primary decisions are displayed for complete decision rules (see Multimedia Appendix 3 and Multimedia Appendix 6). #(X) indicates the count of the last 4 daily check-ins satisfying the condition. Sleep Less-Severe indicates ≤4 hours of sleep. Sleep More-Severe indicates ≥12 hours of sleep or ≥personalized goal upper limit plus 4 hours, whichever is less. Sleep Less indicates <personalized goal lower limit (eg, 6 hours). Sleep More indicates >personalized goal upper limit (eg, 8 hours). Bedtime Window is the personalized 1.5-hour window for going to bed (eg, 10:30 PM to midnight). Risetime Window is the personalized 1.5-hour window for getting up to start the day (eg 7:00 to 8:30 AM). CS: clinical status; |WR|: absolute value of wellness rating.




In addition to varying the content of the *daily review*, the LiveWell expert system also utilizes separate decision tables to prompt users to communicate with their psychiatrists by phone and to deliver emails to activate coaches and enrolled mental health providers to reach out to users via phone (Multimedia Appendix 3 and Multimedia Appendix 7). For *daily check-in* data, a hierarchical decision table organizes selection of clinical reach out messages based on the presence of crisis wellness ratings (+4 or −4), transition to unwell or prodromal states, worsening symptoms when not unwell, or more severe problems with medication adherence and sleep duration (Multimedia Appendix 7). The *daily check-in* clinical reach out decision table thresholds are primarily derived from the psychiatrist survey feedback. A separate set of decision rules trigger clinical reach outs after the *weekly check-in* if the user's score indicates new onset of a manic or depressive episode (Multimedia Appendix 7). The thresholds for the *weekly check-in* are set based on the published thresholds indicating the possible presence of manic or depressive episodes for the ASRM and PHQ8 scales [61,62]. The psychiatrist survey feedback indicated that most psychiatrists felt a higher threshold might be used (Multimedia Appendix 2), but the published thresholds were selected based on the request of the Northwestern University Institutional Review Board. In a report generated for coaches each day (Multimedia Appendix 8), an additional set of decision rules also flags data indicating problems with *daily* or *weekly check-in* adherence (Multimedia Appendix 7).

## Intervention Options

Intervention options include variation in intervention frequency, mode, or content implemented at the decision points following the decision rules [49,51,52]. For LiveWell, the *daily review* provides the primary user interface and explanatory system delivering adaptive content to the user. The *daily review* utilizes a fixed format of 3 to 6 app display pages to adaptively present content for the 26 categories determined by the *daily review* decision rules (Multimedia Appendix 3 and Multimedia Appendix 6) [57]. The first page of the *daily review* provides



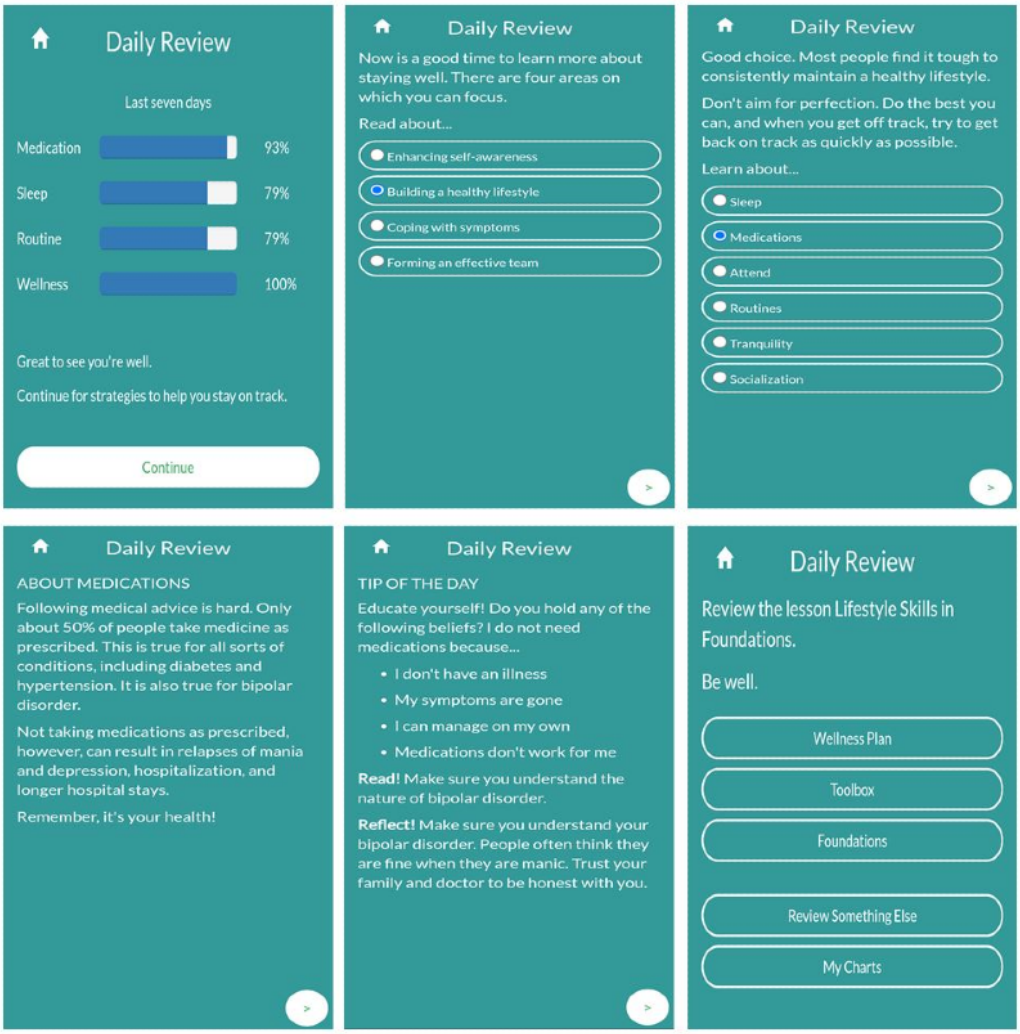
XSL•FO
RenderX

brief feedback about the content category being addressed and the user data indicating why that content category is deemed relevant (Figure 4). In addition, the first page of the *daily review* provides visual feedback regarding target goal achievement using bar graphs to display the percent of days over the last 7 days that the user met the goal for each target. The *daily review* then delivers additional pages of content addressing the selected category with 1 to 4 pages presented depending on the category (Multimedia Appendix 3). The final page of the *daily review* then reiterates the content category being addressed and provides suggestions and links to other sections of the app for review of the selected content category (Figure 4).

To reduce the potential for user fatigue when viewing the same content category multiple times, the first and last pages of the *daily review* randomly present one of a number of pages providing similar content stated in varying ways (Figures 4 and 5). In addition, for pages 2 to 5 of the *daily review*, a given page number (eg, page 2) may have multiple permutations of similar content or multiple unique pages of content that are randomly selected for presentation each time a user views a content category [57]. In some cases, a unique page of content (page 2) is linked to a subsequent unique page of content (page 3) so that only the prior page is randomly selected, but in other cases, subsequent unique pages are randomly selected. Randomly combining permuted and unique pages of content addressing a selected category allows the *daily review* to vary provided content about a given category over multiple views.

The content of each page is typically varied by providing information or suggesting tools to address different motivational, volitional, environmental, or capability-based determinants relevant to the selected category (Figure 5). To provide additional personalization with the aim of maintaining user engagement, some of the *daily review* content categories also present pages where users can select choice options which then determine the content of the subsequent *daily review* page. For instance, if the low-risk category is selected by the inference engine, then page 2 of the *daily review* allows the user to learn more about enhancing self-awareness, building a healthy lifestyle, coping with symptoms, or forming an effective team to assist with staying well (Figure 5). If the user selects lifestyle on page 2, then page 3 allows the user to select 1 of 6 categories (sleep, medication, attend, routine, tranquil, and social) to learn more about supporting different aspects of a healthy lifestyle while living with bipolar disorder.

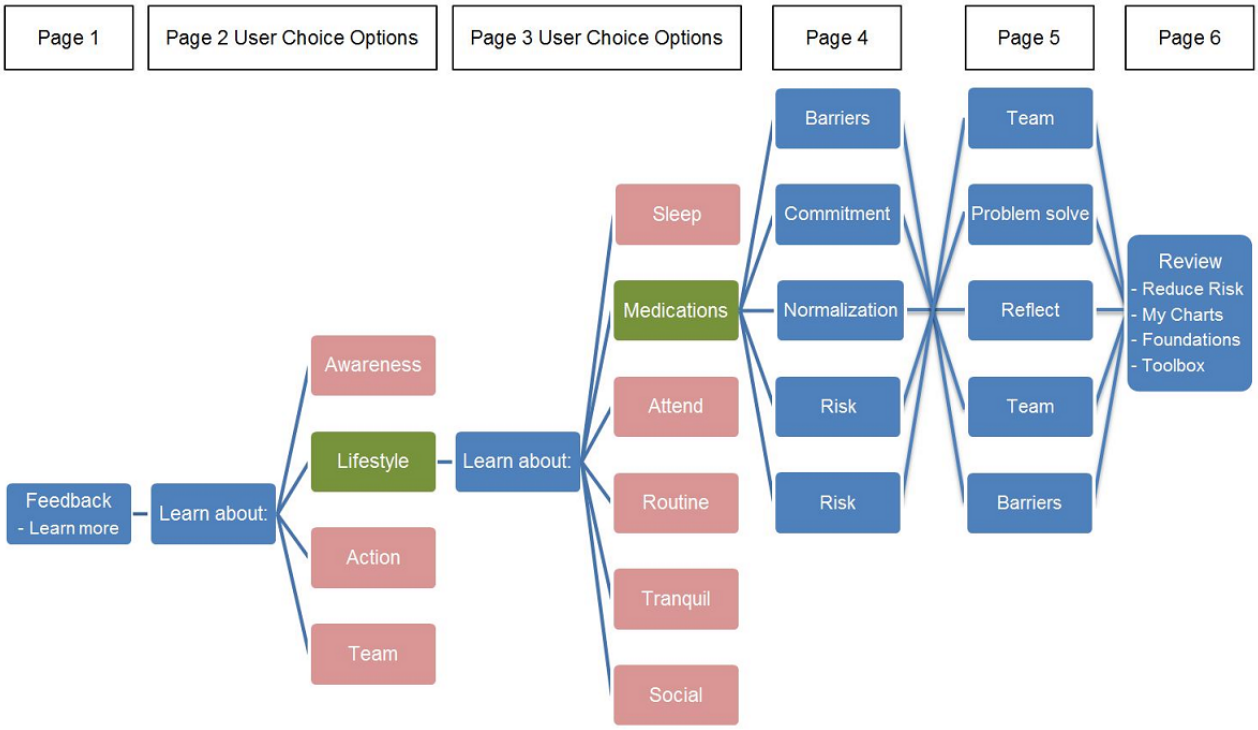
**Figure 4.** Daily review example for the low-risk content category.






XSL•FO
RenderX

**Figure 5.** Daily review flowsheet for the low-risk content category.



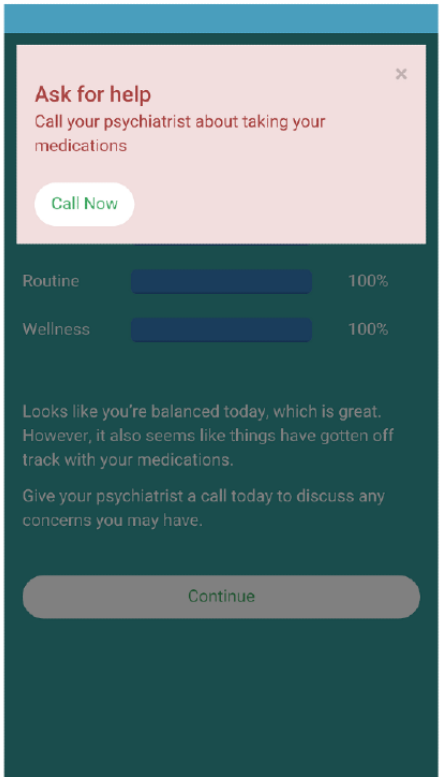

Although 26 content categories are adaptively delivered via the *daily review*, the content as a whole can be grouped into the following 3 tiers: (1) low, moderate, and high risk categories where the user's clinical status is well and early warning signs or worsening symptoms are not endorsed; (2) categories where the user's clinical status is well and early warning signs or worsening symptoms are endorsed or where the user's clinical status is prodromal or recovering; and (3) categories where the user is in an episode or crisis (Figure 3). The first-tier categories contain most of the *daily review* content (71% of available pages), and their content addresses the targets sleep duration, medication adherence, managing signs and symptoms, building support, and keeping a regular routine (30%, 22%, 20%, 9%, and 8% of tier content, respectively) and the following determinants underlying engagement in these target behaviors: attitudes and perceptions, knowledge, support, evaluation, and planning (18%, 15%, 13%, 11%, and 10% of tier content, respectively). For the second-tier categories, the *daily review* content (22% of available pages) addresses the targets managing signs and symptoms and using supports (86% and 6% of tier content, respectively) and the following determinants: evaluation, skills, knowledge, adjustment, and practice (22%, 16%, 16%, 10%, and 7% of tier content, respectively). For the third-tier categories, the *daily review* content (7% of available pages) only addresses managing signs and symptoms (99% of tier content) and the following determinants: evaluation, skills, adjustment, knowledge, and support (22%, 18%, 16%, 15%, and 13% of tier content, respectively). Thus, as users move up the tiers to more acute content categories, the focus of the content shifts to managing signs and symptoms and utilizing supports (Multimedia Appendix 3). In fact, the crisis categories triggered by user entry of a wellness rating of +4 or −4 are the only categories that do not follow the general format of *daily review* content delivery. For these 2 categories, a single page of content without bar graphs is presented requesting that users take immediate action and contact their psychiatrists and supports, and if in danger of self-harm, call 911 or go to the nearest emergency room.

The expert system also prompts clinical care communication to reinforce the importance of engaging supports when having difficulties (Multimedia Appendix 3 and Multimedia Appendix 7). When these decision rules activate a clinical reach out, a pop-up is provided to the user asking the user to call their psychiatrist to address the selected category. To facilitate the user making this call, the pop-up includes a button linked to the psychiatrist's phone number (Figure 6). In addition, an email is sent to the coach and enrolled providers alerting them to the situation (Multimedia Appendix 3). The coach acts on these alerts following structured protocols described in detail elsewhere [58]. Finally, when user adherence with completing *daily* and *weekly check-ins* falls below predetermined thresholds (Multimedia Appendix 7), coaches are also prompted via flagged data in the daily coaching reports (Multimedia Appendix 8). Coaches then follow-up with phone calls, texts, or emails to address adherence issues using a motivational interviewing approach [58]. Overall, the clinical reach out rules aim to shift the mode of intervention delivery from self-management using individual skills to engaging in self-management via effective communication and elicitation of support from mental health care providers and other supports.



XSL•FO
RenderX

**Figure 6.** Daily review pop-up.




## Discussion

The expert system for adaptively delivering content and facilitating clinical care communication for LiveWell is described in detail in this paper to enable replication and ongoing improvement of adaptive MHTs. The LiveWell intervention aims to assist individuals with bipolar disorder in using self-management strategies to manage targets proposed to reduce relapse risk and symptom burden, and improve quality of life [59]. Information from empirically supported psychotherapies for bipolar disorder, health psychology behavior change theories, and chronic disease self-management models directed the selection of the tailoring variables, creation of intervention options, and development of decision rules [59]. User-centered design information guided the timing of the decision points [59]. Feedback from an electronic survey of psychiatrists was combined with existing bipolar disorder literature to define the values of the tailoring variable used in the decision rules.

Because adjunctive psychotherapies for bipolar disorder encourage individuals to manage multiple targets, it was also necessary to develop a hierarchy of tailoring variables ordered as follows: (1) managing signs and symptoms, (2) medication adherence, (3) sleep duration, and (4) routine. A simplified clinical status state machine was developed to predict clinical status based on self-reported wellness rating data. Clinical status from this state machine was then incorporated into hierarchal decision tables (if-then/elseif-then) that select content for users and alerts for providers. For each target behavior selected, intervention options address a variety of determinants proposed to govern engagement in target behaviors, including attitudes and perceptions, knowledge, support, evaluation, planning, skills, adjustment, and practice [57]. When providing adaptive content, the user interface (*daily review*) randomly varies information regarding different determinants and provides user choice options with the goal of reducing user fatigue when category content is delivered multiple times.

There are limitations to the expert system currently developed for the LiveWell intervention. While simplifying the clinical status state machine by using a 9-point wellness rating scale reduces the burden on the user in completing the *daily check-in*, the relationship between the clinical status states determined using these data and those determined by DSM4/CMF outcome assessments has not yet been identified. In addition, while feedback from providers was used to establish the thresholds for adaptive delivery of intervention content, there was significant variation in provider responses related to clinical status, medication adherence, and sleep duration. In their comments, providers also noted that these thresholds may vary between individuals and over time, and that the importance of different targets may also vary for different individuals (Multimedia Appendix 2). Additional data will thus be needed to assess how to optimize the hierarchical organization of feedback delivery and the thresholds for delivering this feedback. Despite these limitations, we hope that the comprehensive description of the expert system delivering the adaptive content for LiveWell and the underlying design decisions addressed during the development of this system will facilitate the ability to replicate, improve, implement, and disseminate effective adaptive MHTs for bipolar disorder and other mental health conditions.



XSL•FO
RenderX

## Conflicts of Interest

EHG has accepted honoraria from Otsuka Pharmaceuticals. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
LiveWell psychiatrist survey.
[PDF File (Adobe PDF File), 67 KB - formative_v5i12e32932_app1.pdf ]

Multimedia Appendix 2
LiveWell psychiatrist survey data.
[XLSX File (Microsoft Excel File), 32 KB - formative_v5i12e32932_app2.xlsx ]

Multimedia Appendix 3
LiveWell daily review logic and content.
[XLSX File (Microsoft Excel File), 187 KB - formative_v5i12e32932_app3.xlsx ]

Multimedia Appendix 4
Dashboard report examples.
[PDF File (Adobe PDF File), 786 KB - formative_v5i12e32932_app4.pdf ]

Multimedia Appendix 5
Psychiatrist survey summary.
[PDF File (Adobe PDF File), 68 KB - formative_v5i12e32932_app5.pdf ]

Multimedia Appendix 6
Daily review decision rules.
[PDF File (Adobe PDF File), 67 KB - formative_v5i12e32932_app6.pdf ]

Multimedia Appendix 7
Clinical and adherence reach out decision rules.
[PDF File (Adobe PDF File), 76 KB - formative_v5i12e32932_app7.pdf ]

Multimedia Appendix 8
Coach report examples.
[XLSX File (Microsoft Excel File), 44 KB - formative_v5i12e32932_app8.xlsx ]

## References

1. American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
2. Vieta E, Berk M, Schulze TG, Carvalho AF, Suppes T, Calabrese JR, et al. Bipolar disorders. Nat Rev Dis Primers 2018 Mar 08;4:18008. [doi: 10.1038/nrdp.2018.8] [Medline: 29516993]
3. Geddes JR, Miklowitz DJ. Treatment of bipolar disorder. The Lancet 2013 May 11;381(9878):1672-1682 [FREE Full text] [doi: 10.1016/S0140-6736(13)60857-0] [Medline: 23663953]
4. Miura T, Noma H, Furukawa TA, Mitsuyasu H, Tanaka S, Stockton S, et al. Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: a systematic review and network meta-analysis. The Lancet Psychiatry 2014 Oct;1(5):351-359. [doi: 10.1016/s2215-0366(14)70314-1]
5. Judd LL, Akiskal HS, Schettler PJ, Endicott J, Leon AC, Solomon DA, et al. Psychosocial disability in the course of bipolar I and II disorders: a prospective, comparative, longitudinal study. Arch Gen Psychiatry 2005 Dec 01;62(12):1322-1330. [doi: 10.1001/archpsyc.62.12.1322] [Medline: 16330720]
6. Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. Arch Gen Psychiatry 2002 Jun;59(6):530-537. [doi: 10.1001/archpsyc.59.6.530] [Medline: 12044195]
7. Judd LL, Schettler PJ, Akiskal HS, Coryell W, Leon AC, Maser JD, et al. Residual symptom recovery from major affective episodes in bipolar disorders and rapid episode relapse/recurrence. Arch Gen Psychiatry 2008 Apr;65(4):386-394. [doi: 10.1001/archpsyc.65.4.386] [Medline: 18391127]



XSL•FO
RenderX

8. Crowe M, Whitehead L, Wilson L, Carlyle D, O'Brien A, Inder M, et al. Disorder-specific psychosocial interventions for bipolar disorder--a systematic review of the evidence for mental health nursing practice. Int J Nurs Stud 2010 Jul;47(7):896-908. [doi: 10.1016/j.ijnurstu.2010.02.012] [Medline: 20233615]
9. Miklowitz DJ. Adjunctive psychotherapy for bipolar disorder: state of the evidence. Am J Psychiatry 2008 Nov;165(11):1408-1419 [FREE Full text] [doi: 10.1176/appi.ajp.2008.08040488] [Medline: 18794208]
10. Schöttle D, Huber CG, Bock T, Meyer TD. Psychotherapy for bipolar disorder: a review of the most recent studies. Curr Opin Psychiatry 2011 Nov;24(6):549-555. [doi: 10.1097/YCO.0b013e32834b7c5f] [Medline: 21918448]
11. Lam DH, Burbeck R, Wright K, Pilling S. Psychological therapies in bipolar disorder: the effect of illness history on relapse prevention - a systematic review. Bipolar Disord 2009 Aug;11(5):474-482. [doi: 10.1111/j.1399-5618.2009.00724.x] [Medline: 19624386]
12. Scott J, Colom F, Vieta E. A meta-analysis of relapse rates with adjunctive psychological therapies compared to usual psychiatric treatment for bipolar disorders. Int J Neuropsychopharmacol 2007 Feb;10(1):123-129 [FREE Full text] [doi: 10.1017/S1461145706006900] [Medline: 16787554]
13. Bond K, Anderson IM. Psychoeducation for relapse prevention in bipolar disorder: a systematic review of efficacy in randomized controlled trials. Bipolar Disord 2015 Jun;17(4):349-362. [doi: 10.1111/bdi.12287] [Medline: 25594775]
14. Gregory VL. Cognitive-behavioral therapy for depression in bipolar disorder: a meta-analysis. J Evid Based Soc Work 2010 Jul;7(4):269-279. [doi: 10.1080/15433710903176088] [Medline: 20799127]
15. Gregory VL. Cognitive-Behavioral Therapy for Mania: A Meta-Analysis of Randomized Controlled Trials. Social Work in Mental Health 2010 Oct 07;8(6):483-494. [doi: 10.1080/15332981003744388]
16. Miklowitz DJ, Efthimiou O, Furukawa TA, Scott J, McLaren R, Geddes JR, et al. Adjunctive Psychotherapy for Bipolar Disorder: A Systematic Review and Component Network Meta-analysis. JAMA Psychiatry 2021 Feb 01;78(2):141-150 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.2993] [Medline: 33052390]
17. Yatham LN, Kennedy SH, Parikh SV, Schaffer A, Bond DJ, Frey BN, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. Bipolar Disord 2018 Mar;20(2):97-170 [FREE Full text] [doi: 10.1111/bdi.12609] [Medline: 29536616]
18. Fountoulakis KN, Grunze H, Vieta E, Young A, Yatham L, Blier P, et al. The International College of Neuro-Psychopharmacology (CINP) Treatment Guidelines for Bipolar Disorder in Adults (CINP-BD-2017), Part 3: The Clinical Guidelines. Int J Neuropsychopharmacol 2017 Feb 01;20(2):180-195 [FREE Full text] [doi: 10.1093/ijnp/pyw109] [Medline: 27941079]
19. Goodwin GM, Haddad PM, Ferrier IN, Aronson JK, Barnes T, Cipriani A, et al. Evidence-based guidelines for treating bipolar disorder: Revised third edition recommendations from the British Association for Psychopharmacology. J Psychopharmacol 2016 Jun;30(6):495-553 [FREE Full text] [doi: 10.1177/0269881116636545] [Medline: 26979387]
20. Olfson M, Marcus SC, Druss B, Pincus HA. National trends in the use of outpatient psychotherapy. Am J Psychiatry 2002 Nov;159(11):1914-1920. [doi: 10.1176/appi.ajp.159.11.1914] [Medline: 12411228]
21. Busch AB, Ling D, Frank RG, Greenfield SF. Changes in the quality of care for bipolar I disorder during the 1990s. Psychiatr Serv 2007 Jan;58(1):27-33 [FREE Full text] [doi: 10.1176/ps.2007.58.1.27] [Medline: 17215409]
22. Brown JD, Barrett A, Hourihan K, Caffery E, Ireys HT. State Variation in the Delivery of Comprehensive Services for Medicaid Beneficiaries with Schizophrenia and Bipolar Disorder. Community Ment Health J 2015 Jul;51(5):523-534. [doi: 10.1007/s10597-015-9857-5] [Medline: 25786723]
23. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):629-640. [doi: 10.1001/archpsyc.62.6.629] [Medline: 15939840]
24. Reinares M, Sánchez-Moreno J, Fountoulakis KN. Psychosocial interventions in bipolar disorder: what, for whom, and when. J Affect Disord 2014 Mar;156:46-55. [doi: 10.1016/j.jad.2013.12.017] [Medline: 24439829]
25. Denizard-Thompson NM, Feiereisel KB, Stevens SF, Miller DP, Wofford JL. The digital divide at an urban community health center: implications for quality improvement and health care access. J Community Health 2011 Jun;36(3):456-460. [doi: 10.1007/s10900-010-9327-5] [Medline: 21086028]
26. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
27. Ben-Zeev D, Davis KE, Kaiser S, Krzsos I, Drake RE. Mobile technologies among people with serious mental illness: opportunities for future services. Adm Policy Ment Health 2013 Jul;40(4):340-343 [FREE Full text] [doi: 10.1007/s10488-012-0424-x] [Medline: 22648635]
28. Torous J, Wisniewski H, Liu G, Keshavan M. Mental Health Mobile Phone App Usage, Concerns, and Benefits Among Psychiatric Outpatients: Comparative Survey Study. JMIR Ment Health 2018 Nov 16;5(4):e11715 [FREE Full text] [doi: 10.2196/11715] [Medline: 30446484]
29. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr 02;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]




XSL•FO
RenderX

30. Michalak EE, Suto MJ, Barnes SJ, Hou S, Lapsley S, Scott MW, Crest Bd, et al. Effective self-management strategies for bipolar disorder: A community-engaged Delphi Consensus Consultation study. J Affect Disord 2016 Dec;206:77-86 [FREE Full text] [doi: 10.1016/j.jad.2016.06.057] [Medline: 27466745]
31. Murray G, Suto M, Hole R, Hale S, Amari E, Michalak EE. Self-management strategies used by 'high functioning' individuals with bipolar disorder: from research to clinical practice. Clin Psychol Psychother 2011 Apr;18(2):95-109. [doi: 10.1002/cpp.710] [Medline: 20572206]
32. Janney CA, Bauer MS, Kilbourne AM. Self-management and bipolar disorder--a clinician's guide to the literature 2011-2014. Curr Psychiatry Rep 2014 Sep;16(9):485. [doi: 10.1007/s11920-014-0485-5] [Medline: 25123130]
33. Russell SJ, Browne JL. Staying well with bipolar disorder. Aust N Z J Psychiatry 2005 Mar 26;39(3):187-193. [doi: 10.1080/j.1440-1614.2005.01542.x] [Medline: 15701069]
34. Nicholas J, Boydell K, Christensen H. Beyond symptom monitoring: Consumer needs for bipolar disorder self-management using smartphones. Eur Psychiatry 2017 Jul;44:210-216. [doi: 10.1016/j.eurpsy.2017.05.023] [Medline: 28692910]
35. Daus H, Kislicyn N, Heuer S, Backenstrass M. Disease management apps and technical assistance systems for bipolar disorder: Investigating the patients´ point of view. J Affect Disord 2018 Mar 15;229:351-357. [doi: 10.1016/j.jad.2017.12.059] [Medline: 29331693]
36. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, et al. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. J Am Med Inform Assoc 2016 May;23(3):477-484. [doi: 10.1093/jamia/ocv165] [Medline: 26911822]
37. Todd NJ, Jones SH, Lobban FA. What do service users with bipolar disorder want from a web-based self-management intervention? A qualitative focus group study. Clin Psychol Psychother 2013;20(6):531-543. [doi: 10.1002/cpp.1804] [Medline: 22715161]
38. Kazdin AE. Mediators and mechanisms of change in psychotherapy research. Annu Rev Clin Psychol 2007;3:1-27. [doi: 10.1146/annurev.clinpsy.3.022806.091432] [Medline: 17716046]
39. Kazdin AE, Nock MK. Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. J Child Psychol Psychiatry 2003 Nov;44(8):1116-1129. [doi: 10.1111/1469-7610.00195] [Medline: 14626454]
40. Nock MK. Conceptual and design essentials for evaluating mechanisms of change. Alcohol Clin Exp Res 2007 Oct;31(10 Suppl):4s-12s. [doi: 10.1111/j.1530-0277.2007.00488.x] [Medline: 17880341]
41. Laurenceau J, Hayes AM, Feldman GC. Some methodological and statistical issues in the study of change processes in psychotherapy. Clin Psychol Rev 2007 Jul;27(6):682-695 [FREE Full text] [doi: 10.1016/j.cpr.2007.01.007] [Medline: 17328996]
42. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
43. Clough BA, Casey LM. Technological adjuncts to enhance current psychotherapy practices: a review. Clin Psychol Rev 2011 Apr;31(3):279-292. [doi: 10.1016/j.cpr.2010.12.008] [Medline: 21382535]
44. Lindhiem O, Bennett CB, Rosen D, Silk J. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. Behav Modif 2015 Nov;39(6):785-804 [FREE Full text] [doi: 10.1177/0145445515595198] [Medline: 26187164]
45. Areàn PA, Hoa Ly K, Andersson G. Mobile technology for mental health assessment. Dialogues Clin Neurosci 2016 Jun;18(2):163-169 [FREE Full text] [Medline: 27489456]
46. Broekhuizen K, Kroeze W, van Poppel MNM, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286 [FREE Full text] [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
47. Lustria MLA, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
48. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013 Sep;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
49. Collins LM, Murphy SA, Bierman KL. A conceptual framework for adaptive preventive interventions. Prev Sci 2004 Sep;5(3):185-196 [FREE Full text] [doi: 10.1023/b:prev.0000037641.26017.00] [Medline: 15470938]
50. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
51. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]



XSL•FO
RenderX

52. Malmartel A, Ravaud P, Ghosn L, Tran V. A classification of methods used to personalize participative interventions revealed inadequate reporting in trial protocols. J Clin Epidemiol 2021 May;133:80-93. [doi: 10.1016/j.jclinepi.2021.01.004] [Medline: 33476767]
53. Brendryen H, Kraft P, Schaalma H. Looking Inside the Black Box: Using Intervention Mapping to Describe the Development of the Automated Smoking Cessation Intervention 'Happy Ending'. J Smok Cessat 2012 Feb 21;5(1):29-56. [doi: 10.1375/jsc.5.1.29]
54. Brendryen H, Johansen A, Nesvåg S, Kok G, Duckert F. Constructing a Theory- and Evidence-Based Treatment Rationale for Complex eHealth Interventions: Development of an Online Alcohol Intervention Using an Intervention Mapping Approach. JMIR Res Protoc 2013 Jan 23;2(1):e6 [FREE Full text] [doi: 10.2196/resprot.2371] [Medline: 23612478]
55. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
56. Spring B. Evidence-based practice in clinical psychology: what it is, why it matters; what you need to know. J Clin Psychol 2007 Jul;63(7):611-631. [doi: 10.1002/jclp.20373] [Medline: 17551934]
57. Goulding EH, Dopke C, Rossom R, Michaels T, Martin C, Ryan C, et al. Empirical and Theoretical Framework, Intervention Design, and Study Protocol for a Randomized Controlled Trial of LiveWell: A Smartphone-based Self-management Intervention for Individuals with Bipolar Disorder. JMIR Research Protocols (forthcoming) 2021. [doi: 10.2196/30710]
58. Dopke CA, McBride A, Babington P, Jonathan GK, Michaels T, Ryan C, et al. Development of Coaching Support for LiveWell: A Smartphone-Based Self-Management Intervention for Bipolar Disorder. JMIR Form Res 2021 Mar 24;5(3):e25810 [FREE Full text] [doi: 10.2196/25810] [Medline: 33759798]
59. Jonathan GK, Dopke CA, Michaels T, Bank A, Martin CR, Adhikari K, et al. A Smartphone-Based Self-management Intervention for Bipolar Disorder (LiveWell): User-Centered Development Approach. JMIR Ment Health 2021 Apr 12;8(4):e20424 [FREE Full text] [doi: 10.2196/20424] [Medline: 33843607]
60. Jonathan GK, Dopke CA, Michaels T, Martin CR, Ryan C, McBride A, et al. A Smartphone-Based Self-Management Intervention for Individuals with Bipolar Disorder (LiveWell): Qualitative Study on User Experiences of the Behavior Change Process. JMIR Ment Health 2021 Nov 22;8(11):e32306 [FREE Full text] [doi: 10.2196/32306] [Medline: 34813488]
61. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
62. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. Biological Psychiatry 1997 Nov 15;42(10):948-955. [doi: 10.1016/S0006-3223(96)00548-3] [Medline: 9359982]
63. Havermans R, Nicolson NA, Devries MW. Daily hassles, uplifts, and time use in individuals with bipolar disorder in remission. J Nerv Ment Dis 2007 Sep;195(9):745-751. [doi: 10.1097/NMD.0b013e318142cbf0] [Medline: 17984774]
64. Miklowitz DJ, Goodwin GM, Bauer MS, Geddes JR. Common and specific elements of psychosocial treatments for bipolar disorder: a survey of clinicians participating in randomized trials. J Psychiatr Pract 2008 Mar;14(2):77-85 [FREE Full text] [doi: 10.1097/01.pra.0000314314.94791.c9] [Medline: 18360193]
65. Miklowitz DJ, Scott J. Psychosocial treatments for bipolar disorder: cost-effectiveness, mediating mechanisms, and future directions. Bipolar Disord 2009 Jun;11 Suppl 2:110-122. [doi: 10.1111/j.1399-5618.2009.00715.x] [Medline: 19538690]
66. Lam DH, Bright J, Jones S, Hayward P, Schuck N, Chisholm D, et al. Cognitive Therapy for Bipolar Illness—A Pilot Study of Relapse Prevention. Cognitive Therapy and Research 2000;24(5):503-520. [doi: 10.1023/A:1005557911051]
67. Goossens PJJ, Kupka RW, Beentjes TAA, van Achterberg T. Recognising prodromes of manic or depressive recurrence in outpatients with bipolar disorder: a cross-sectional study. Int J Nurs Stud 2010 Oct;47(10):1201-1207. [doi: 10.1016/j.ijnurstu.2010.01.010] [Medline: 20189172]
68. Keitner GI, Solomon DA, Ryan CE, Miller IW, Mallinger A, Kupfer DJ, et al. Prodromal and residual symptoms in bipolar I disorder. Comprehensive Psychiatry 1996 Sep;37(5):362-367. [doi: 10.1016/s0010-440x(96)90018-8]
69. Wong G, Lam D. The development and validation of the coping inventory for prodromes of mania. Journal of Affective Disorders 1999 Apr;53(1):57-65. [doi: 10.1016/s0165-0327(98)00096-2] [Medline: 10363667]
70. Sachs GS, Guille C, McMurrich SL. A clinical monitoring form for mood disorders. Bipolar Disord 2002 Oct;4(5):323-327. [doi: 10.1034/j.1399-5618.2002.01195.x] [Medline: 12479665]
71. Sachs GS, Thase ME, Otto MW, Bauer M, Miklowitz D, Wisniewski SR, et al. Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD). Biological Psychiatry 2003 Jun 01;53(11):1028-1042. [doi: 10.1016/s0006-3223(03)00165-3] [Medline: 12788248]
72. De Dios C, Ezquiaga E, Agud JL, Vieta E, Soler B, García-López A. Subthreshold symptoms and time to relapse/recurrence in a community cohort of bipolar disorder outpatients. J Affect Disord 2012 Dec 20;143(1-3):160-165. [doi: 10.1016/j.jad.2012.05.047] [Medline: 22925351]
73. Gitlin MJ, Swendsen J, Heller TL, Hammen C. Relapse and impairment in bipolar disorder. Am J Psychiatry 1995 Nov;152(11):1635-1640. [doi: 10.1176/ajp.152.11.1635] [Medline: 7485627]
74. Perlis RH, Ostacher MJ, Patel JK, Marangell LB, Zhang H, Wisniewski SR, et al. Predictors of recurrence in bipolar disorder: primary outcomes from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). Am J Psychiatry 2006 Feb;163(2):217-224. [doi: 10.1176/appi.ajp.163.2.217] [Medline: 16449474]




XSL•FO
RenderX

## Abbreviations

**ASRM:** Altman Self-Rating Mania
**CMF:** clinical monitoring form
**DSM4:** Diagnostic and Statistical Manual of Mental Disorders fourth edition
**MHT:** mental health technology
**PHQ8:** Patient Health Questionnaire-8

*Edited by G Eysenbach; submitted 15.08.21; peer-reviewed by J Zulueta, P Li; comments to author 24.09.21; revised version received 23.10.21; accepted 28.10.21; published 24.12.21.*

*Please cite as:*
*Goulding EH, Dopke CA, Michaels T, Martin CR, Khiani MA, Garborg C, Karr C, Begale M*
*A Smartphone-Based Self-management Intervention for Individuals With Bipolar Disorder (LiveWell): Protocol Development for an Expert System to Provide Adaptive User Feedback*
*JMIR Form Res 2021;5(12):e32932*
*URL: https://formative.jmir.org/2021/12/e32932*
*doi:10.2196/32932*
*PMID:34951598*

©Evan H Goulding, Cynthia A Dopke, Tania Michaels, Clair R Martin, Monika A Khiani, Christopher Garborg, Chris Karr, Mark Begale. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Detection of Clenbuterol-Induced Changes in Heart Rate Using At-Home Recorded Smartwatch Data: Randomized Controlled Trial

Willem O Elzinga[1], MSc; Samantha Prins[1,2], MSc; Laura G J M Borghans[1], PhD; Pim Gal[1,2], MD, PhD; Gabriel A Vargas[3], MD, PhD; Geert J Groeneveld[1,2], MD, PhD; Robert J Doll[1], PhD

[1]Centre for Human Drug Research, Leiden, Netherlands
[2]Leiden University Medical Center, Leiden, Netherlands
[3]CuraSen Therapeutics Inc, San Carlos, CA, United States

**Corresponding Author:**
Robert J Doll, PhD
Centre for Human Drug Research
Zernikedreef 8
Leiden, 2333CL
Netherlands
Phone: 31 715246400
Email: rjdoll@chdr.nl

## Abstract

**Background:** Although electrocardiography is the gold standard for heart rate (HR) recording in clinical trials, the increasing availability of smartwatch-based HR monitors opens up possibilities for drug development studies. Smartwatches allow for inexpensive, unobtrusive, and continuous HR estimation for potential detection of treatment effects outside the clinic, during daily life.

**Objective:** The aim of this study is to evaluate the repeatability and sensitivity of smartwatch-based HR estimates collected during a randomized clinical trial.

**Methods:** The data were collected as part of a multiple-dose, investigator-blinded, randomized, placebo-controlled, parallel-group study of 12 patients with Parkinson disease. After a 6-day baseline period, 4 and 8 patients were treated for 7 days with an ascending dose of placebo and clenbuterol, respectively. Throughout the study, the smartwatch provided HR and sleep state estimates. The HR estimates were quantified as the 2.5th, 50th, and 97.5th percentiles within awake and asleep segments. Linear mixed models were used to calculate the following: (1) the intraclass correlation coefficient (ICC) of estimated sleep durations, (2) the ICC and minimum detectable effect (MDE) of the HR estimates, and (3) the effect sizes of the HR estimates.

**Results:** Sleep duration was moderately repeatable (ICC=0.64) and was not significantly affected by study day ($P$=.83), clenbuterol ($P$=.43), and study day by clenbuterol ($P$=.73). Clenbuterol-induced changes were detected in the asleep HR as of the first night (+3.79 beats per minute [bpm], $P$=.04) and in the awake HR as of the third day (+8.79 bpm, $P$=.001). The median HR while asleep had the highest repeatability (ICC=0.70). The MDE (N=12) was found to be smaller when patients were asleep (6.8 bpm to 11.7 bpm) than while awake (10.7 bpm to 22.1 bpm). Overall, the effect sizes for clenbuterol-induced changes were higher while asleep (0.49 to 2.75) than while awake (0.08 to 1.94).

**Conclusions:** We demonstrated the feasibility of using smartwatch-based HR estimates to detect clenbuterol-induced changes during clinical trials. The asleep HR estimates were most repeatable and sensitive to treatment effects. We conclude that smartwatch-based HR estimates obtained during daily living in a clinical trial can be used to detect and track treatment effects.

**Trial Registration:** Netherlands Trials Register NL8002; https://www.trialregister.nl/trial/8002

***(JMIR Form Res 2021;5(12):e31890)*** doi:10.2196/31890

**KEYWORDS**

photoplethysmography; smartwatch; wearable; at-home; heart rate; RCT; wearable device; digital health; cardiovascular; cardiology; sensors; heart rate sensor; smart technology



XSL•FO
RenderX

## Introduction

In early phase clinical trials, heart rhythm is one of the vital signs that is routinely measured with laboratory grade equipment (eg, electrocardiogram [ECG]) for safety and to detect adverse treatment effects [1]. Repeated recording of heart rhythm parameters, such as the heart rate (HR), provides clinically relevant information on the effects of a novel drug intervention. The monitoring of the HR aids in the detection of early signs of distress [2], for example, pharmacologically induced arrhythmias, hypotension, or gastrointestinal symptoms. Currently, there is a trend towards at-home observation of patients enrolled in clinical trials [3]. Recording of the HR in a home environment would require a system that is capable of continuous recording and is relatively user-friendly.

One of the strategies to obtain at-home recordings of the HR is to use a Holter monitor. This requires a setup in which a small device and several electrodes are attached to a patient, which could make this device cumbersome to use. Various smaller wearable devices are now widely available which include ECG patches (eg, [4,5]) or photoplethysmography (PPG) sensors integrated in wrist-worn devices [5]. While continuous monitoring using an ECG patch can be accurate and sensitive to medication (eg, [6]), wrist-worn PPG sensors are more convenient to use. Additionally, wrist-worn HR monitors using PPG sensors are relatively cheap commercial devices and may provide clinically relevant information in addition to being user-friendly. Furthermore, PPG HR estimates have shown to correlate well with ECG HR estimates in healthy participants [7,8]. The PPG sensor measures volumetric changes in blood flow within the microvasculature during the cardiac cycle [9]. In clinical trials, such monitors could be used to unobtrusively obtain estimates of the HR at home (ie, away from the clinic) and during daily life activities (eg, while asleep or awake). Therefore, if HR monitoring for a longer duration is required, smartwatch HR monitors could potentially result in shorter or fewer clinic visits, reducing the total trial costs.

Before using such a device as a biomarker source in clinical trials, the biomarker correlation with the actual HR, its repeatability, and its sensitivity need to be evaluated. HR estimates from PPG wearables have been demonstrated to correlate well with the actual HR [7,8]. Next, biomarkers must be able to produce repeatable values in the presence of measurement noise within the subject. Finally, the sensitivity of the biomarker to detect meaningful changes must be demonstrated. Before these qualities can be assessed, however, the continuous stream of HR estimates must be preprocessed into 1 or several biomarkers.

HR estimates obtained from a wearable device during physical activity may show higher variability and are less accurate than estimates obtained while at rest [10]. As a result, it is useful to distinguish between the active and rest HR estimates prior to further analysis. A relatively simple method to distinguish between rest and active estimates is by splitting the HR estimates into two groups: HR estimates while asleep and HR estimates while awake. As an alternative to asking patients to keep track of their sleep pattern by means of a diary, wearable devices can also provide information on sleep patterns (eg, sleep duration, light or deep sleep). However, while the estimated sleep duration can be considered reasonably reliable, estimates of the substates (eg, light or deep sleep) cannot [11]. Therefore, the estimated sleep start and end times can be used to split the HR stream into awake and asleep estimates.

Here, we evaluate the feasibility of using a smartwatch-based HR monitor in a clinical trial. We first evaluate the repeatability of the estimated sleep durations. Then, after evaluating the repeatability of smartwatch-obtained HR over a time period of 1 week, we present the sensitivity of the smartwatch to detect drug-induced changes in the HR. We used smartwatch-based HR monitoring in a trial of patients with Parkinson disease (PD) who were treated with clenbuterol, a $\beta_2$-adrenoceptor agonist, in a placebo-controlled fashion. The work presented here was part of a trial aimed at demonstrating the central nervous system effects of the drug. The results of that trial will be reported separately.

## Methods

### Study Location and Ethical Approval

The work presented here is a part of an exploratory objective of a larger clinical trial. The trial was approved by the Stichting Beoordeling Ethiek Biomedisch Onderzoek ethics committee, Assen, the Netherlands, and was executed in accordance with the Declaration of Helsinki at the Centre for Human Drug Research, Leiden, the Netherlands. The trial is registered in the Netherlands Trial Register under NL8002. Here, we report the exploratory objective of evaluating the effects of clenbuterol, a direct-acting sympathomimetic with predominant $\beta_2$-adrenoceptor selectivity, on HR measured with a smartwatch.

### Participants

All participants provided written informed consent. A total of 12 patients with PD (10 men, 2 women; mean age 64.3 years, SD 6.6 years) participated in the trial. The trial included patients at a Hoehn and Yahr stage between 1 and 3 and a Mini Mental Status Examination score of above 25. Patients taking β-blockers or β-agonists for reasons other than for treatment of tremor were not eligible. Other medication for PD (eg, levodopa, carbidopa, and benserazide) had to be at a stable dosage for at least 3 months. Patients were otherwise healthy or stable (eg, stable hypertension for at least 3 months, stable dyslipidemia, normal or not clinically significant 12-lead ECG, normal laboratory values including hematology and chemistry values). The use of vitamin E (up to 400 IU daily), estrogens, aspirin (81 to 300 mg daily), blood pressure medications (except for adrenergic agents), and cholesterol-lowering agents was allowed if treatment was stable for 3 months prior to screening.

### Study Design

This was a multiple-dose, investigator-blinded, randomized, placebo-controlled, parallel group study. The study duration was approximately 7 weeks, including a screening period of up to 31 days, but a minimum of 6 days (day –6 to day 0), a treatment period of 1 week (day 0 to day 6), and a 1-week



XSL•FO
RenderX

follow-up period. Patients started wearing the smartwatch after the screening session.

Patients were randomly assigned to either the placebo or clenbuterol treatment group with a 1:2 ratio (ie, 4 patients were assigned to the placebo group and 8 patients were assigned to the clenbuterol group). Patients continued taking their own medication throughout the full study period. The clenbuterol dose was titrated from 20 µg on the first treatment day (ie, day 0) to 40 µg on day 1 and 80 µg on days 2 through 6. Patients were dosed with either clenbuterol or a matching number of tablets of placebo. Since the tablets of clenbuterol and placebo were nonmatching, patients were instructed not to discuss the tablet appearance with other patients or study site staff.

## Measurements

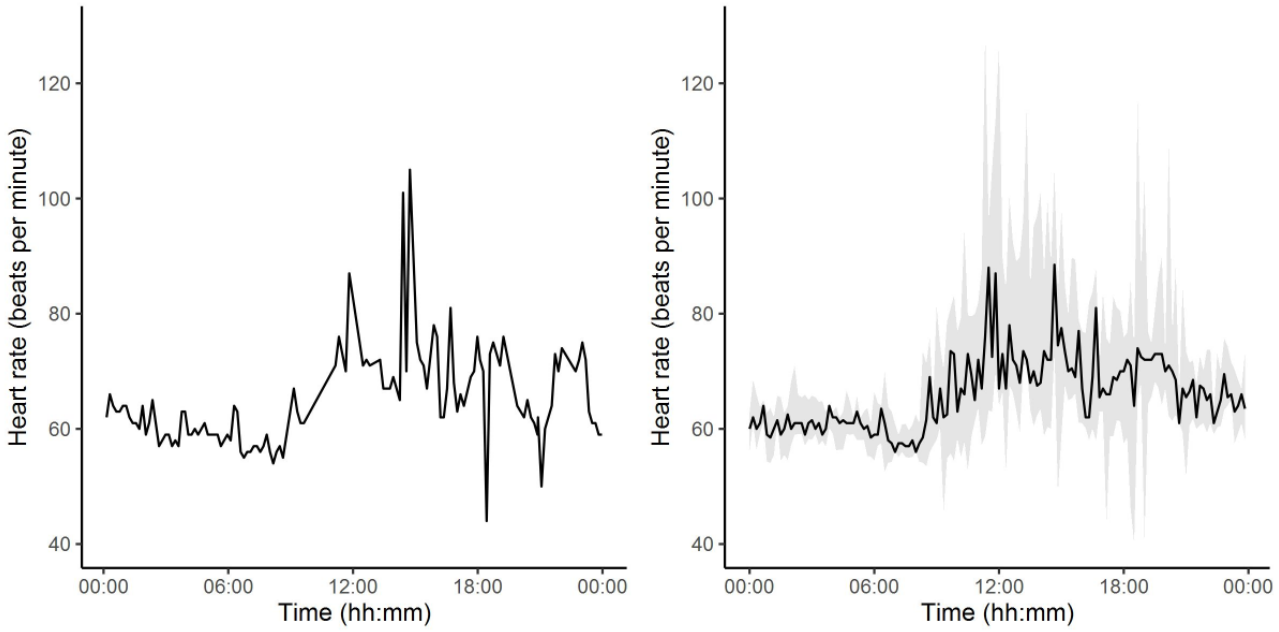
Measurements were collected using a smartwatch (Steel HR, Withings). During the screening session, patients were given a smartwatch and the Withings Health Mate app was installed on their phones. The app was configured to work with the smartwatch and connected to an account that was used to upload the measurements during the study days. Patients wore the smartwatch at least 6 days prior to the treatment period (ie, day –6 to day 0) and continued wearing the smartwatch during the treatment period (ie, day 0 to day 6). As the quality of the measurements is potentially variable [10], participants were instructed to position and strap the watch such that motion on the skin was minimized throughout the study period.

Data collected by the smartwatch include estimates of the HR and sleep states. The smartwatch provides an HR estimate every 10 minutes. However, when physical activity (eg, running) is detected, the sample frequency increases. The smartwatch provided 3 discrete states for sleep: awake, light sleep, or deep sleep. Only changes in states were stored by the smartwatch. All HR samples and changes in sleep state samples contained timestamps.

## Smartwatch Data Preprocessing

The smartwatch continuously captured HR for 1 week prior to treatment and 1 week after treatment (ie, day –6 to day 6). An example HR time series of 1 patient is presented in Figure 1. The left and right panel present the HR throughout a single day and the median HR throughout the baseline period, with the shading representing the 2.5th and 97.5th HR percentiles, respectively.

**Figure 1.** (A) Smartwatch-obtained heart rate (HR) throughout a single day in a single subject. (B) Median heart rate throughout the day for a single subject, with shading between the 2.5th and 97.5th HR percentiles.




For further processing steps, we distinguished between the HR while asleep and the HR while awake. The HR while asleep is defined as all HR samples collected between the first moment the smartwatch indicated a sleep state until the last sleep state change of the following day. The HR while awake is defined as all HR samples collected outside the asleep boundaries. Asleep HR series shorter than 3 hours and nights without any detected sleep state were excluded from further analysis. For days with missing sleep states, the awake period was assumed to start at 9 AM and end at 9 PM. For each asleep or awake period and for each patient, HR series were summarized by their 2.5th, 50th, and 97.5th percentile values. From here on, these markers will be referred to as Awake-Low, Awake-Median, Awake-High, Asleep-Low, Asleep-Median, and Asleep-High.

## Statistical Analysis

All data preparation and analyses were performed using the R software package (R Foundation for Statistical Computing). The linear mixed models (LMM) described below were fitted using the nlme package. Restricted maximum likelihood was used for fitting and the Kenward-Roger method was used for estimating the degrees of freedom. Contrasts between treatment days and effect sizes were estimated using the emmeans package. Statistical significance was defined as $P<.05$.



XSL•FO
RenderX

### Repeatability of Smartwatch Sleep Duration Estimates

As HR series might be more variable while asleep than while awake (as found in [10]), the HR series were segmented into asleep and awake periods using the smartwatch-provided sleep states. To assess the reliability of patients' sleep states, an LMM was fit to estimate if treatment affected the total sleep duration. The intercept, day (ie, day –6 to 6), treatment (ie, placebo or clenbuterol), as well as the interaction between day and treatment were included as fixed effects. Between-subject random effects were only included for the intercept. Type III *F* statistics were used to assess if treatment significantly explained the sleep duration variability. The reliability of sleep duration was quantified as the intraclass correlation coefficient (ICC), which is defined as the ratio between the between-subject variance and the sum of the between-subject and within-subject variance.

### Repeatability and Minimum Detectable Effect of Smartwatch HR Estimates

To assess the repeatability and the minimum detectable effect (MDE) of HR estimates of the markers, only the predose data were considered. An LMM was fitted for each of the 6 markers. The intercept and day (ie, day –6 to day 0) were included as fixed effects. Between-subject random effects were only included for the intercept. The type III *F* statistic of the fixed effect for day was used to assess if structural day-to-day variability was present. The repeatability of each marker was quantified as the ICC.

The MDE values were calculated using an inverse sample size (n=12) calculation. For each marker, the total SD was calculated by taking the square root of the sum of within-subject and between-subject variation. The SD was then used in the inverse sample size calculation together with quantiles from the *t*-distribution with n–1 degrees of freedom for a false positive probability α of .05 and a power of .80.

### Effect of Clenbuterol on the Smartwatch HR Estimates

To estimate the effect of the treatment on the smartwatch-obtained HR, an LMM was fit for each of the smartwatch markers. The intercept, day (ie, day –6 to 6), treatment (ie, placebo or clenbuterol), as well as the interaction between day and treatment were included as fixed effects. Between-subject random effects were only included for the intercept. Type III *F* statistics were used to assess the main and interaction effects. The estimated effects on each dosing day (day 0 to day 6) were contrasted to the average predose (day –6 to day –1) values. Bonferroni correction was applied to correct for multiple testing. The Cohen *d* effect size was calculated as the contrast between HR estimates after clenbuterol treatment and the averaged predose HR, divided by the square root of the sum of the between-subject and within-subject variation.

## Results

All participants completed the experiment procedures. None of the patients were taking β-blockers or β-agonists. There were, on average, 116 awake and 39 asleep HR estimates available per patient per day. Due to the late inclusion of 1 patient, 5 days of baseline data were missing for this patient. Another patient reported sleepless nights due to shoulder pain, resulting in 10 missing nights for this patient. Another total of 14 nights were missing for the other patients due to missing sleep states (n=12) or nights lasting shorter than 3 hours (n=2). For the patients in the clenbuterol group, 4 out of 44 nights were missing prior to dosing and 6 out of 48 nights were missing post dosing.

### Repeatability of Smartwatch Sleep Duration Estimates

The day by treatment interaction effect did not have a significant effect on the total sleep duration ($F_{12,92.5}$=0.72, *P*=.73). The sleep duration was also not significantly affected by the main effects day ($F_{12,92.5}$=0.60, *P*=.83) and treatment ($F_{1,9.2}$=0.69, *P*=.43). The repeatability of the sleep duration was moderate (ICC=0.64).

### Repeatability and MDE of Smartwatch HR Estimates

For all 6 repeatability models, no significant effect of day was detected on the smartwatch HR (Table 1). The repeatability values are presented in Table 2. Except for the Awake-High HR (ICC=0.23), all markers showed moderate repeatability values with ICC values ranging between 0.58 and 0.70. The Asleep-Median marker showed the best repeatability (ICC=0.70). The MDE values (both absolute and relative) are presented in Table 2. The asleep HR markers had lower MDE values than the awake HR markers. The Asleep-Low and Awake-High markers had the best and worst MDE values, respectively.

**Table 1.** Type III *F* statistics for each of the 6 models used to assess the repeatability of the smartwatch HR (heart rate). Each row depicts the values for the day effect.

| Model | Degrees of freedom, numerator/denominator | *F* statistic | *P* value |
|---|---|---|---|
| **Awake** | | | |
| Low | 5/50.5 | 1.44 | .23 |
| Median | 5/50.4 | 0.86 | .52 |
| High | 5/50.8 | 0.28 | .92 |
| **Asleep** | | | |
| Low | 5/40.9 | 1.93 | .11 |
| Median | 5/40.7 | 1.25 | .30 |
| High | 5/40.8 | 1.14 | .35 |



XSL•FO
RenderX

**Table 2.** The estimated mean predose heart rate (HR), standard error (SE), intraclass correlation coefficient (ICC), and minimum detectable effect (MDE) for each heart rate (HR) marker. The MDE percentage is relative to the estimated mean HR.

| HR marker | Mean (SE), bpm[a] | ICC | MDE, bpm (%) |
|---|---|---|---|
| **Asleep** | | | |
| Low | 57.6 (1.70) | 0.64 | 6.84 (11.7) |
| Median | 63.2 (1.36) | 0.70 | 7.79 (12.2) |
| High | 74.2 (1.94) | 0.67 | 11.71 (15.1) |
| **Awake** | | | |
| Low | 58.2 (2.03) | 0.58 | 10.72 (18.3) |
| Median | 75.4 (1.56) | 0.65 | 11.45 (15.5) |
| High | 104.1 (2.55) | 0.23 | 22.13 (21.5) |

[a]bpm: beats per minute.

## Effect of Clenbuterol on Smartwatch HR Estimates

Table 3 presents the results of the 6 models used to assess the effect of clenbuterol on each of the smartwatch HR markers. The day by treatment effect is significant for all markers, except for the Awake-High marker. Post hoc comparisons between the average baseline HR and HR during treatment days are presented in Table 4 and in Figure 2. Note that the Awake-High values are not displayed in Table 4 as the day by treatment interaction term was not significant for this marker. No changes in HR were detected for any of the markers in the placebo group. In the clenbuterol group, a change in HR was detected by the Asleep-Low HR marker as of the first night after treatment (+3.79 beats per minute [bpm], $P$=.04). A change in HR was detected by the other asleep HR markers as of the second night (Asleep-Median: +7.9 bpm, $P$<.001; Asleep-High: +9.5 bpm, $P$<.001). For the Awake HR markers, a change in HR was detected by the Awake-Low HR marker as of the third day after treatment (+8.79 bpm, $P$=.001). Table 5 presents the effect sizes to detect changes induced by clenbuterol. The effect sizes are generally better for the asleep HR markers than for the awake HR markers. Additionally, effect sizes increase with increasing dosages of clenbuterol.

**Table 3.** Type III $F$ statistics for each of 6 models used to assess the treatment effect on the smartwatch-obtained heart rate.

| Model | Low | | Median | | High | |
|---|---|---|---|---|---|---|
| | $F$ (df1/df2)[a] | $P$ value | $F$ (df1/df2) | $P$ value | $F$ (df1/df2) | $P$ value |
| **Awake** | | | | | | |
| Day | 3.05 (12/115.0) | .001 | 5.07 (12/115.0) | <.001 | 1.12 (12/115.0) | .35 |
| Treatment | 0.47 (1/10.0) | .51 | 0.09 (1/10.3) | .76 | 0.25 (1/10.0) | .63 |
| Day × Treatment | 3.52 (12/115.0) | <.001 | 5.74 (12/115.0) | <.001 | 0.44 (12/115.0) | .94 |
| **Asleep** | | | | | | |
| Day | 11.94 (12/91.3) | <.001 | 10.67 (12/91.3) | <.001 | 3.70 (12/91.3) | <.001 |
| Treatment | 1.27 (1/10.0) | .29 | 0.50 (1/10.0) | .50 | 0.03 (1/10.0) | .87 |
| Day × Treatment | 10.20 (12/91.3) | <.001 | 11.38 (12/91.3) | <.001 | 4.84 (12/91.3) | <.001 |

[a]Df1/df2 indicate the numerator/denominator degrees of freedom.

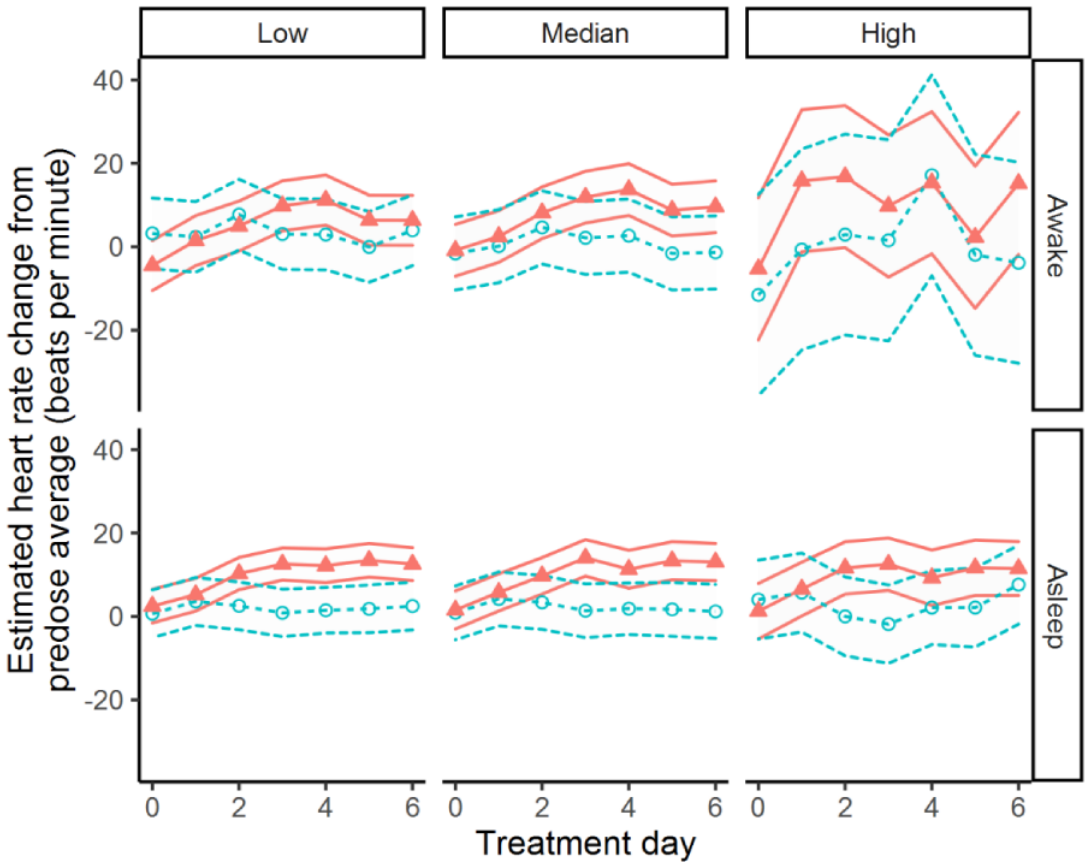


XSL•FO
RenderX

**Table 4.** Post hoc comparisons between postdose and averaged predose heart rates (mean difference ± SE). The *P* values after Bonferroni correction are provided.

| | Asleep | | | | | | Awake | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Day | Low | *P* value | Median | *P* value | High | *P* value | Low | *P* value | Median | *P* value |
| **Clenbuterol** | | | | | | | | | | |
| 0 | 3.79 (1.34) | .04 | 3.7 (1.43) | .08 | 4.21 (2.42) | .60 | –0.68 (2.26) | .99 | 1.79 (1.92) | .99 |
| 1 | 6.54 (1.25) | <.001 | 7.90 (1.33) | <.001 | 9.50 (2.26) | <.001 | 5.30 (2.26) | .15 | 5.10 (1.92) | .06 |
| 2 | 11.60 (1.18 | <.001 | 11.87 (1.26) | <.001 | 14.57 (2.13) | <.001 | 8.79 (2.26) | .001 | 10.79 (1.92) | <.001 |
| 3 | 13.88 (1.18) | <.001 | 16.18 (1.26) | <.001 | 15.47 (2.13) | <.001 | 13.64 (2.26) | <.001 | 14.54 (1.92) | <.001 |
| 4 | 13.49 (1.34) | <.001 | 13.45 (1.43) | <.001 | 12.23 (2.42) | <.001 | 15.01 (2.26) | <.001 | 16.35 (1.92) | <.001 |
| 5 | 14.77 (1.34) | <.001 | 15.48 (1.43) | <.001 | 14.62 (2.43) | <.001 | 10.18 (2.26) | <.001 | 11.41 (1.92) | <.001 |
| 6 | 13.85 (1.25) | <.001 | 15.18 (1.33) | <.001 | 14.43 (2.26) | <.001 | 10.16 (2.26) | <.001 | 12.22 (1.92) | <.001 |
| **Placebo** | | | | | | | | | | |
| 0 | –0.66 (1.86) | .99 | –1.27 (1.99) | .99 | 1.08 (3.37) | .99 | –0.61 (3.15) | .99 | –4.23 (2.67) | .82 |
| 1 | 2.27 (1.86) | .99 | 2.06 (1.99) | .99 | 2.77 (3.37) | .99 | –1.45 (3.15) | .99 | –2.48 (2.67) | .99 |
| 2 | 1.23 (1.86) | .99 | 1.23 (1.99) | .99 | –2.93 (3.37) | .99 | 3.90 (3.15) | .99 | 2.02 (2.67) | .99 |
| 3 | –0.46 (1.92) | .99 | –0.77 (2.05) | .99 | –4.80 (3.48) | .99 | –0.76 (3.15) | .99 | –0.48 (2.67) | .99 |
| 4 | 0.16 (1.67) | .99 | –0.28 (1.78) | .99 | –0.82 (3.02) | .99 | –0.88 (3.15) | .99 | 0.02 (2.67) | .99 |
| 5 | 0.51 (1.86) | .99 | –0.44 (1.99) | .99 | –0.81 (3.37) | .99 | –3.86 (3.15) | .99 | –4.23 (2.67) | .82 |
| 6 | 1.17 (1.86) | .99 | –0.94 (1.99) | .99 | 4.68 (3.37) | .99 | 0.14 (3.15) | .99 | –3.98 (2.67) | .98 |

**Figure 2.** Estimated daily heart rate (HR) changes from average baseline values. The blue circles and orange triangles indicate the HR in the placebo arm and the clenbuterol arm, respectively. The 95% confidence intervals of the estimated HR are indicated by the lines with treatment arm color and style.






XSL•FO
RenderX

**Table 5.** Cohen *d* effect sizes to detect clenbuterol treatment.

| | Asleep, *d* effect size | | | Awake, *d* effect size | | |
|---|---|---|---|---|---|---|
| Day | Low | Median | High | Low | Median | High |
| 0 | 0.73 | 0.63 | 0.49 | –0.08 | 0.21 | –0.20 |
| 1 | 1.25 | 1.34 | 1.11 | 0.64 | 0.60 | 0.68 |
| 2 | 2.22 | 2.01 | 1.70 | 1.06 | 1.28 | 0.72 |
| 3 | 2.66 | 2.75 | 1.80 | 1.64 | 1.72 | 0.42 |
| 4 | 2.58 | 2.28 | 1.42 | 1.81 | 1.94 | 0.66 |
| 5 | 2.83 | 2.63 | 1.70 | 1.22 | 1.35 | 0.11 |
| 6 | 2.65 | 2.58 | 1.68 | 1.22 | 1.45 | 0.65 |

# Discussion

## Principal Findings

HR estimates from wearable PPG sensors have been demonstrated to correlate well with the actual HR [8,10]. Here, we further explored the feasibility of using a smartwatch-based PPG HR monitor in clinical trials by assessing the repeatability and sensitivity of HR estimates. All data presented here were part of a randomized placebo-controlled trial including patients with Parkinson disease. Patients wore a smartwatch for a total duration of 13 days. HR estimates were separated into asleep and awake periods based on the smartwatch provided sleep states After the first 6 days, patients started taking daily doses of either a placebo or clenbuterol. Clenbuterol is a $\beta_2$-adrenoceptor agonist known to increase HR [12]. Clenbuterol has been used clinically for decades and has a relatively benign safety profile. The main effects are cardiovascular, with acute increases in HR, increases in systolic blood pressure, and decreases in diastolic blood pressure. Studies in patients with reversible obstructive airways disease treated with clenbuterol have demonstrated statistically significant increases in HR [13], and case reports in humans mention tachycardia as an adverse effect [13-15].

While the smartwatch estimated sleep and awake stages, it is not clear how exactly the algorithm derives these states. The sleep duration can be estimated reasonably well using consumer smartwatches in healthy participants [11]. However, it is unknown how well this algorithm performs in a patient population and under treatment conditions. As the HR and movement is affected by clenbuterol, this could also act as a confounding parameter in the sleep state algorithm. We found, however, that the sleep duration was estimated with moderate repeatability (ICC=0.64) and was not significantly affected by the treatment. Additionally, the number of missing nights was similar both prior to treatment and during treatment. Therefore, we argue that the sleep stage estimates are fit for our purpose.

The smartwatch HR estimates were also found to be moderately repeatable, but most repeatable when patients were asleep (Table 2). The lower repeatability values for awake HR measures can be explained by the larger within-subject variability during the day. Other research groups also found that HR measures at rest are more repeatable than when active [10]. In addition to the higher repeatability values, asleep HR measures had lower MDE values suggesting higher sensitivity while patients were asleep than while patients were awake. Therefore, HR measures while asleep can be considered more repeatable and are potentially more sensitive to detecting changes in HR than HR measures while awake.

Both awake and asleep HR measures were able to detect the effect of clenbuterol (Table 3). In line with the MDE values, the asleep HR measures showed higher sensitivity than the awake HR measures. The increase in HR due to clenbuterol was detected on the first day after dosing using the asleep HR measures but was detected only after the first postdose day using the awake HR measures (Table 4). Additionally, the Awake-High HR marker was unable to detect a change at all due to the large within-subject and between-subject variability (Figure 2). On the other hand, the Asleep-Low HR marker showed the highest sensitivity. All in all, the asleep HR measures are more likely to detect medication effects than awake HR measures.

Drug development trials are generally expensive [15], especially when increasing the number of participants and trial duration (eg, phase 2 or 3 trials). Continuous HR and sleep state values are typically unavailable in clinical trials, especially when subjects are ambulant. However, smartwatches can easily be integrated in such trials for continuous monitoring of participants who go about their daily routines. Additionally, combining a smartwatch with a remote monitoring app installed on a mobile phone (eg, to track mobile phone sensor data) allows passive tracking of both social and physical activities. It should be noted, however, that many smartwatches are not marked as a medical device and therefore cannot be used for diagnostic or treatment purposes. Regardless, smartwatches can detect treatment-induced changes in the HR.

## Limitations

A limitation of this study is that although the asleep HR measures showed higher sensitivity to clenbuterol-induced changes than the awake HR measures, the number of missing data points was substantially higher during the night than during the day. The data loss might be explained by the following: (1) PD-related sleep disturbances [16], (2) reduced compliance with the study protocol, or (3) the use of estimated sleep states provided by the smartwatch. Patients might find the smartwatch unpleasant to wear during the night and decide to either loosen the strap or remove the watch from the wrist. Additionally, at



XSL•FO
RenderX

least 1 patient mentioned having trouble falling asleep and being restless throughout the night. This could result in either sleepless nights or inaccurate estimates of sleep states. Instead of relying on estimated sleep states, a fixed time window could be used (eg, between 11 PM and 6 AM).

Another limitation is the arbitrary choice for the quantification of the HR time series. Although we demonstrate that these are sensitive to treatment-induced changes in the HR, other percentiles could have been chosen as well. More sophisticated time series analysis could be used to describe the HR patterns (eg, nonlinear regression analysis). The second arbitrary choice we made in this study is that we assumed patients to be awake between 9 AM and 9 PM when sleep states were missing during the night. Such a choice can be improved by, for example, adding a pedometer or other sensor to detect steps or movements.

## Conclusion

We demonstrated the feasibility of using smartwatch-based HR measurement in early phase clinical trials. The smartwatch was able to detect drug-induced changes in the HR without the need for patients to be present in-clinic. Additionally, such devices allow researchers to obtain frequently sampled HR estimates with limited discomfort for the patient. We demonstrated that HR estimates captured while asleep were more repeatable and most sensitive to drug-induced changes than measures taken while awake. Future studies for validation of commercially available devices should include larger sample sizes and could also include other relevant cardiac parameters, such as the HR variability.

## Acknowledgments

This study was carried out at the Centre for Human Drug Research in Leiden, the Netherlands, and was sponsored by CuraSen Therapeutics, Inc in San Carlos, California. We would like to thank George Sangster for his constructive feedback during the writing of the manuscript.

## Conflicts of Interest

GAV is the Chief Medical Officer of CuraSen Therapeutics, Inc and owns shares in this company.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 792 KB - formative_v5i12e31890_app1.pdf ]

## References

1. Dias D, Paulo Silva Cunha J. Wearable health devices-vital sign monitoring, systems and technologies. Sensors (Basel) 2018 Jul 25;18(8):2414 [FREE Full text] [doi: 10.3390/s18082414] [Medline: 30044415]
2. Tisdale JE, Chung MK, Campbell KB, Hammadah M, Joglar JA, Leclerc J, et al. Drug-induced arrhythmias: a scientific statement from the American Heart Association. Circulation 2020 Oct 13;142(15):e214-e233. [doi: 10.1161/cir.0000000000000905]
3. Downey C, Soon S, Jayne DG, Savadottir H. Wearable devices for remote vital signs monitoring in the outpatient setting: an overview of the field. BMJ Innov 2020 Jan 14;6(2):55-71. [doi: 10.1136/bmjinnov-2019-000354]
4. Torfs T, Smeets CJP, Geng D, Berset T, Van der Auwera J, Vandervoort P, et al. Clinical validation of a low-power and wearable ECG patch for long term full-disclosure monitoring. J Electrocardiol 2014 Nov;47(6):881-889. [doi: 10.1016/j.jelectrocard.2014.08.012] [Medline: 25201418]
5. Biswas D, Simoes-Capela N, Van Hoof C, Van Helleputte N. Heart rate estimation from wrist-worn photoplethysmography: a review. IEEE Sens J 2019 Aug 15;19(16):6560-6570. [doi: 10.1109/jsen.2019.2914166]
6. Izmailova ES, McLean IL, Hather G, Merberg D, Homsy J, Cantor M, et al. Continuous monitoring using a wearable device detects activity-induced heart rate changes after administration of amphetamine. Clin Transl Sci 2019 Nov;12(6):677-686 [FREE Full text] [doi: 10.1111/cts.12673] [Medline: 31365190]
7. Weiler DT, Villajuan SO, Edkins L, Cleary S, Saleem JJ. Wearable heart rate monitor technology accuracy in research: a comparative study between PPG and ECG technology. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Thousand Oaks, CA: SAGE Publishing; 2017 Oct 20 Presented at: Human Factors and Ergonomics Society Annual Meeting; October 9-13, 2017; Austin, TX p. 1292-1296. [doi: 10.1177/1541931213601804]
8. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016 Sep 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
9. Lemay M, Bertschi M, Sola J, Renevey P, Parak J, Korhonen I. Application of optical heart rate monitoring. In: Sazonov E, Neuman MR, editors. Wearable Sensors: Fundamentals, Implementation and Applications, 1st Edition. Cambridge, MA: Academic Press; 2014:129.
10. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020 Feb 10;3(1):1-9 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]



XSL•FO
RenderX

11. Mantua J, Gravel N, Spencer RMC. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 May 05;16(5):1-11 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
12. Spiller HA, James KJ, Scholzen S, Borys DJ. A descriptive study of adverse events from clenbuterol misuse and abuse for weight loss and bodybuilding. Subst Abus 2013 Jul;34(3):306-312. [doi: 10.1080/08897077.2013.772083]
13. Whitsett T, Manion C, Wilson M. Cardiac, pulmonary and neuromuscular effects of clenbuterol and terbutaline compared with placebo. Br J Clin Pharmacol 1981 Aug 26;12(2):195-200 [FREE Full text] [doi: 10.1111/j.1365-2125.1981.tb01200.x] [Medline: 7306434]
14. Brett J, Dawson AH, Brown JA. Clenbuterol toxicity: a NSW poisons information centre experience. Med J Aust 2014 Mar 03;200(4):219-221. [doi: 10.5694/mja13.10982] [Medline: 24580525]
15. Rawlins MD. Cutting the cost of drug development? Nat Rev Drug Discov 2004 Apr;3(4):360-364. [doi: 10.1038/nrd1347] [Medline: 15060531]
16. Comella CL. Sleep disorders in Parkinson's disease: an overview. Mov Disord 2007 Sep;22 Suppl 17:S367-S373 [FREE Full text] [doi: 10.1002/mds.21682] [Medline: 18175398]

## Abbreviations

**ECG:** electrocardiogram
**HR:** heart rate
**ICC:** intraclass correlation coefficient
**LMM:** linear mixed model
**MDE:** minimum detectable effect
**PPG:** photoplethysmography

*Edited by G Eysenbach; submitted 09.07.21; peer-reviewed by IV George, A Mariakakis; comments to author 30.07.21; revised version received 12.08.21; accepted 21.11.21; published 30.12.21.*

*Please cite as:*
*Elzinga WO, Prins S, Borghans LGJM, Gal P, Vargas GA, Groeneveld GJ, Doll RJ*
*Detection of Clenbuterol-Induced Changes in Heart Rate Using At-Home Recorded Smartwatch Data: Randomized Controlled Trial*
*JMIR Form Res 2021;5(12):e31890*
*URL: https://formative.jmir.org/2021/12/e31890*
*doi:10.2196/31890*
*PMID:34967757*

©Willem O Elzinga, Samantha Prins, Laura G J M Borghans, Pim Gal, Gabriel A Vargas, Geert J Groeneveld, Robert J Doll. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Predicting Risk of Stroke From Lab Tests Using Machine Learning Algorithms: Development and Evaluation of Prediction Models

Eman M Alanazi[1,2], MSc; Aalaa Abdou[3], MD; Jake Luo[4], PhD

[1]Department of Health Informatics, College of Health Sciences, Saudi Electronic University, Riyadh, Saudi Arabia
[2]Department of Biomedical and Health Informatics, College of Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI, United States
[3]Department of Radiotherapy, Children's Cancer Hospital 57357, Cairo, Egypt
[4]Department of Health Informatics and Administration, College of Health Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

**Corresponding Author:**
Eman M Alanazi, MSc
Department of Health Informatics
College of Health Sciences
Saudi Electronic University
Abi Bakr As Siddiq Branch Rd
Riyadh, 13323
Saudi Arabia
Phone: 966 112613500
Email: e.alanazi@seu.edu.sa

## Abstract

**Background:** Stroke, a cerebrovascular disease, is one of the major causes of death. It causes significant health and financial burdens for both patients and health care systems. One of the important risk factors for stroke is health-related behavior, which is becoming an increasingly important focus of prevention. Many machine learning models have been built to predict the risk of stroke or to automatically diagnose stroke, using predictors such as lifestyle factors or radiological imaging. However, there have been no models built using data from lab tests.

**Objective:** The aim of this study was to apply computational methods using machine learning techniques to predict stroke from lab test data.

**Methods:** We used the National Health and Nutrition Examination Survey data sets with three different data selection methods (ie, without data resampling, with data imputation, and with data resampling) to develop predictive models. We used four machine learning classifiers and six performance measures to evaluate the performance of the models.

**Results:** We found that accurate and sensitive machine learning models can be created to predict stroke from lab test data. Our results show that the data resampling approach performed the best compared to the other two data selection techniques. Prediction with the random forest algorithm, which was the best algorithm tested, achieved an accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the curve of 0.96, 0.97, 0.96, 0.75, 0.99, and 0.97, respectively, when all of the attributes were used.

**Conclusions:** The predictive model, built using data from lab tests, was easy to use and had high accuracy. In future studies, we aim to use data that reflect different types of stroke and to explore the data to build a prediction model for each type.

***(JMIR Form Res 2021;5(12):e23440)*** doi:10.2196/23440

**KEYWORDS**

stroke; lab tests; machine learning technology; predictive analytics

## Introduction

Stroke is a neurological deficit, primarily because of acute central nervous system focal injury caused by a vascular issue. It is a major cause of disability and death worldwide [1]. It estimated that the overall prevalence of stroke in the United States is 2.5%, and about 7 million Americans over the age of 20 years have experienced a stroke. The condition has a significant negative impact on patients' health and quality of life. It also has a negative impact on hospital services and the availability of beds and was estimated to cost the US economy about US $351.2 billion between 2014 and 2015 [2]. There are



XSL•FO
RenderX

two types of stroke: ischemic and hemorrhagic. Hemorrhagic stroke occurs because of a burst vessel that leads to bleeding in the brain, whereas ischemic stroke occurs because of a blockage of the arteries of the brain. Ischemic strokes are the most common, comprising 85% to 90% of all strokes [3]. This condition can be prevented by promoting health and increasing awareness of risk factors. There are many risk factors related to lifestyle, including obesity, diet, alcohol intake, and lack of physical activity [4]. Underlying conditions, such as diabetes, hypertension, and cardiovascular diseases, may also lead to stroke. Therefore, proper self-management of these diseases and the pursuit of a healthy lifestyle may prevent the occurrence of stroke.

In 2019, the American College of Cardiology/American Heart Association released the Guideline on the Primary Prevention of Cardiovascular Disease. The guideline recommends a complete assessment and examination of patients who are at risk of developing blockages in their arteries that may lead to a heart attack or stroke and might die as a result [5]. Now more than ever, physicians can access clinical evidence to identify high-risk patients using approaches such as acquiring a complete patient history and conducting thorough physical exams for risk assessment. Patient records contain many useful predictive factors, such as patient demographic (eg, age and gender), lifestyle (eg, diet and physical activity), and existing medical condition factors (eg, diabetes and hypertension), that might lead to stroke [5]. The growth of arterial blockages and decades of damage to blood vessels, which may lead to stroke, are often associated with these risk factors. If physicians can assess the risks of stroke easily and conveniently, strokes could be prevented at an earlier stage. This approach could save lives and reduce the economic burden of health care services. In the age of artificial intelligence and machine learning, a clinical decision support system has been developed to assist physicians to diagnose and identify individuals with a high risk of stroke. The potential of applying machine learning technologies in the cardiovascular domain is significant, from identifying individuals with a high risk of stroke [6,7] to predicting outcomes of patients following treatment [8,9]. Most of these studies use either health habits and lifestyle factors, such as smoking or alcohol consumption; conditions that predispose to strokes, such as hypertension and diabetes mellites; or neuroimaging, such as computed tomography and magnetic resonance imaging, to either classify or predict the disease.

Besides assessing known risk factors for stroke, scientists are trying to develop lab tests that can predict stroke. One of the major advantages of using lab test results for prediction is that lab tests are commonly collected in clinical settings, and the information is often well documented in patients' records. In this study, we explored data-driven approaches using supervised machine learning models to predict the risk of stroke from different lab tests.

Several studies have been able to identify independent laboratory tests that are correlated with stroke using descriptive statistical analysis. Sughrue et al [10] conducted a retrospective study in 2013 that identified 35 tests with a statistically significant correlation with a future stroke diagnosis. The most informative were for various types of cholesterol. Two of these 35 laboratory tests were urine tests, and 33 were blood, serum, or plasma tests. Some tests were positively associated with an outcome of stroke (ie, neutrophil count and percent; CD3+, CD8+, and T8 suppressor cells; monocytes; eosinophils; and CD3 cells), and others were negatively correlated (ie, hematocrit and lymphocytes). Their results show that it is possible to correlate future stroke with collected lab test data. Farah and Samra [11] conducted a retrospective study investigating the association between the neutrophil-to-lymphocyte ratio (NLR), mean platelet volume (MPV), and the risk of stroke. Two-tailed *t* tests showed no significant differences in the stroke group's MPV values compared with those in the control group. However, the NLRs of the stroke patients were significantly different compared with those of the control group. That study indicated the existence of a correlation between the level of NLR and stroke risk. NLR levels have been shown to be higher in stroke patients than in control groups. Feng et al [12] reviewed the scientific literature on the potential role and the possible epidemiological relationships between red cell distribution width (RDW) and ischemic stroke in a meta-analysis of 40 manuscripts from China National Knowledge Infrastructure and PubMed databases. They reported that patients with stroke had higher levels of RDW than those without strokes. Another study by Kaya et al [13] also investigated the association between baseline RDW level and stroke risk in patients with heart failure. These authors found that heart failure patients suffering from stroke had significantly increased basal RDW levels (mean 16.9, SD 1.14, vs mean 14.8, SD 1.6; $P<.001$) and serum uric acid levels (mean 8.8, SD 1.7, vs mean 7.5, SD 1.1; $P=.027$) compared with patients without stroke, according to the propensity score analysis. Giles et al [14] used data from a national cohort to investigate whether low folate levels were associated with ischemic stroke and found that folate concentrations of ≤9.2 nmol/L could be a risk factor for ischemic stroke (relative risk 1.37, 95% CI 0.82-2.29). Another study by Qin et al [15] concluded that there is a significant risk of first ischemic stroke in hypertensive patients with low levels of folate and vitamin B12.

These studies demonstrate the value of lab test results for predicting stroke. Our study aimed to leverage lab test results to build machine learning models for stroke prediction. We prepared the data sets using three data selection techniques for this study. After that, for each data selection technique, we applied four individual machine learning classifiers to prepare prediction models. We measured the performance of each prediction model using six different performance measures. Our results indicate that the data resampling technique outperformed the decision tree and random forest classifiers.

## Methods

### Overview

Figure 1 shows the outline of our investigation. In the first step, we collected data from the National Health and Nutrition Examination Survey (NHANES). In the second step, we selected the data using three data techniques for our prediction models. The first one was conducted without data resampling, the second

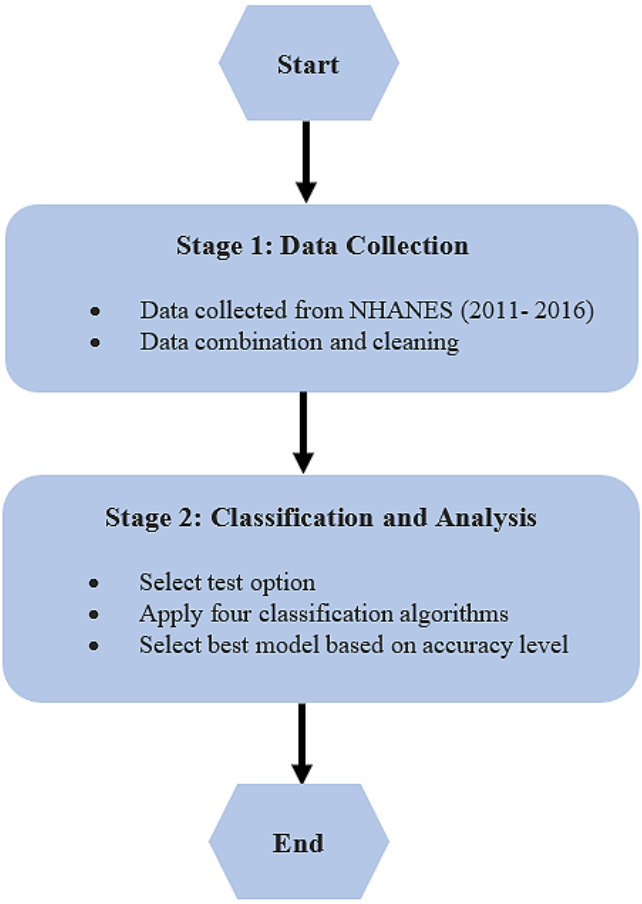


XSL•FO
RenderX

one included data imputation, and the third one was conducted with data resampling.

We used 10-fold cross-validation to perform the train and test approach. To train models, we used four different machine learning classifiers, and six performance measures were used to assess the performance of the models. The elaborated descriptions of the data sets, classifiers, and performance metrics that were used are given below.

**Figure 1.** Flow diagram of the study methodology. NHANES: National Health and Nutrition Examination Survey.




## Data Collection

The NHANES survey was conducted to examine the health and nutritional status of adults and children in the United States; "NHANES is a major program of the National Center for Health Statistics (NCHS). NCHS is part of the Centers for Disease Control and Prevention (CDC) and has the responsibility for producing vital and health statistics for the Nation" [16]. The data sets contain five domains: demographics, dietary data, examination data, laboratory data, and questioner data. Each domain contains several subdomains. Our focus was on data sets that contain information about laboratory tests. The data sets are available from 1999 to 2017, and we considered data from 2011 to 2015. The total number of participants was 15,714 during this period. To reduce the impact of imbalanced data, we noted that in the entire data set, there were about 17% of participants who had experienced a stroke. Therefore, we included total of 4186 participants, of whom 608 (14.5%) had experienced a stroke (Figure 2). The list of data attributes is shown in Table 1. The data sets contained 21 attributes, including each patient's age and gender as well as other lab test information for each respective patient. The data sets and their information are available online [16], where the data are presented from the year 2000 to the current year. For this study, the data were collected for each year and combined using the sequence number (SEQN). After combining and cleaning the data, we used the Waikato Environment for Knowledge Analysis (WEKA; version 3.8.0) system to build and test machine learning models.



XSL•FO
RenderX

**Figure 2.** Participant selection and prevalence of stroke in the National Health and Nutrition Examination Survey (NHANES).

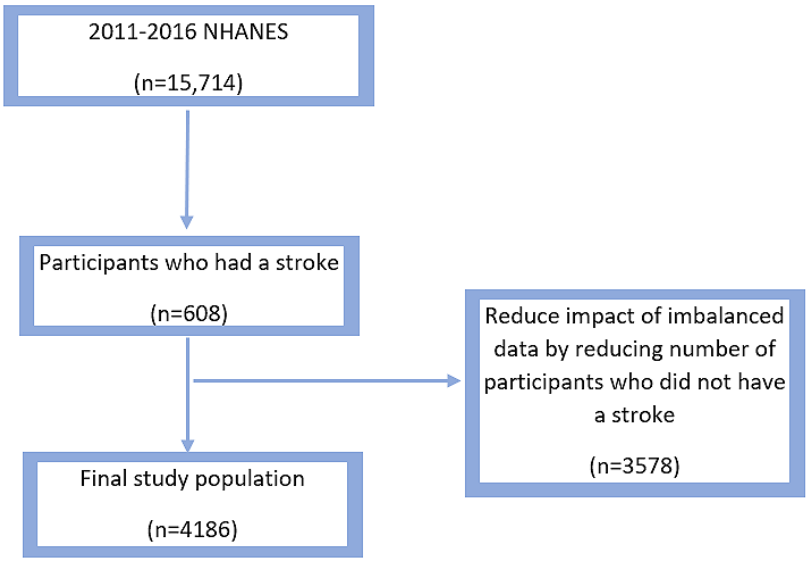



**Table 1.** List of the data attributes.

| Feature[a] | Units |
|---|---|
| Age | Years |
| Gender | N/A[b] |
| Albumin, urine | ug/mL |
| Creatinine, urine | mg/dL |
| White blood cell count | 1000 cells/µL |
| Lymphocytes | 1000 cells/µL |
| Monocytes | 1000 cells/µL |
| Segmented neutrophils | 1000 cells/µL |
| Eosinophils | 1000 cells/µL |
| Basophils | 1000 cells/µL |
| Red blood cell count | Million cells/µL |
| Hemoglobin | g/dL |
| Hematocrit | % |
| Mean cell volume | fL |
| Mean cell hemoglobin | pg |
| Mean corpuscular hemoglobin concentration | g/dL |
| Red cell distribution width | % |
| Platelet count | 1000 cells/µL |
| Mean platelet volume | fL |
| Cotinine, serum | ng/mL |
| Red blood cell folate | mg/dL |

[a]All data types were numeric, except for "gender," which was nominal.
[b]N/A: not applicable; this type of data did not have units.

## Classification

Several different machine learning algorithms can handle a binary classification problem. In this study, we used four machine learning algorithms: naïve Bayes, BayesNet, J48 (Java implementation of C4.5 algorithm), and random forest. The performance of the algorithms was evaluated and compared for



XSL•FO
RenderX

stroke prediction using lab test results as features. Details of the algorithms are as follows:

- The J48 algorithm creates a tree based on the C4.5 algorithm with pruning.
- The random forest algorithm creates a forest of random trees and outputs the mode of the classes created by individual trees.
- The naïve Bayes algorithm creates a classifier based on the naïve Bayes method, which assumes that all attributes are independent.
- The BayesNet algorithm creates a classifier based on non–naïve Bayes, which does not assume that all attributes are independent.

In the cross-validation approach, the data sets are divided into several equal portions; in general, 5-fold and 10-fold cross-validations are used when the data sets are equally divided into 5 and 10 portions [17]. With this approach, for each simulation, one portion of each data set is used to train the prediction model and the rest are used for validation. In this study, we used 10-fold cross-validation and, in this process, we divided the whole of each data set into 10 equal parts; each time, 10% of each data set was used to train the model and 90% was used for validation. In this task, three data analyses were conducted where the first data analysis applied each of the machine learning techniques on the data sets without data manipulation or resampling. The aim was to determine the baseline for the data sets among the various machine learning techniques. The imputation of missing data set entries was conducted in the second analysis. In statistics, imputation entails substituting missing data with values calculated using any of a number of techniques [18]. Imputation is a useful technique in remedying missing data, since missing data may lead to inaccurate predictions. We used the default ReplaceMissingValue filter in WEKA, which replaces all missing values for nominal and numeric attributes in a data set with the modes and means from the training data. Most of the features had 5% missing values, and one feature had 11% missing values. After the imputation of the missing data, data resampling was conducted in the third analysis. Data resampling is a commonly used technique, since training may result in nonuniformity of class labels. In this case, the resampling technique was applied to select a specific subset of data points for model training [19]. After resampling the data, the results of the first analysis should be improved because of the balancing of the data set distribution. A balanced distribution was achieved through the use of WEKA, which randomly resamples the data. Based on the available theoretical knowledge about resampling and imputation in statistics, the results after the third analysis should be improved.

### Evaluation Metrics

Model accuracy was evaluated based on the following measures: recall or sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and area under the curve (AUC) (or area under the receiver operating characteristic [ROC] curve) to compare the four classifiers. Details of these measures are as follows:

- Sensitivity, also known as recall or true positive rate, is the number of true positives divided by the number of true positives plus the number of false negatives. It is the likelihood that the patient has a high risk of stroke [20].
- Specificity, also known as the true negative rate, is the proportion of individuals classified as nonstroke to the total number of actual nonstroke cases. It is the likelihood that a patient who does not have a risk of stroke will have a negative result [21].
- PPV, also known as precision, is the number of true positives divided by the number of true positives plus the number of false positives. It is the proportion of individuals who have suffered a stroke to the total number of participants classified as having a risk of stroke [22].
- NPV is the percentage of negative tests in patients who are free from the disease or the proportion of individuals who have not suffered a stroke to the total number of participants classified as not having a risk of stroke [22].
- Overall accuracy is the number of correctly classified instances over the total size of the data set [20].
- The AUC is the area under the ROC curve, which is constructed by plotting the true positive rate against the true negative rate [23].

We will also look at the Pearson correlation coefficient value of each independent predictor to investigate the relationship between each lab test and risk of stroke.

## Results

In the NHANES data sets, 608 participants suffered from a stroke from 2011 to 2015. The median age of participants who had a stroke was 51 years for both men and women. The numbers of men and women who had a stroke were 220 (36.2%) and 190 (31.3%), respectively; 198 (32.6%) participants did not reveal their gender identity.

After the data collection process, the data were analyzed in three ways: without data resampling, with data imputation, and with data resampling. Data resampling techniques were used to tackle data imbalance problems in the data sets. These sampling techniques are widely used in machine learning–based prediction models in different areas [24]. Our first analysis was done without the data resampling technique, where the four machine learning algorithms were applied directly to the data sets. The first analysis produced poor results for all four classifiers. The best sensitivity rate among the classifiers in the first analysis was for the BayesNet model, followed by the naïve Bayes model. In the second analysis, we applied the data imputation technique to the data sets, which replaced missing values and deleted features that had more than 50% missing values; the prediction accuracy improved for all models, except for the naïve Bayes model, whose performance decreased slightly after replacing the missing values.

In the third analysis, we resampled the data. After resampling, the prediction accuracy improved significantly for both the decision tree and random forest models, but only slightly for the naïve Bayes and BayesNet models. Table 2 shows the scores of accuracy, sensitivity, specificity, PPV, NPV, and AUC, according to the three data analysis techniques and four



XSL•FO
RenderX

classifiers. The table shows that the random forest model was the best classifier with the data resampling technique. Figures 3 and 4 show the score comparisons among the three data selection techniques for the decision tree and random forest models, respectively. We considered the decision tree and random forest classifiers to compare the performance, as they significantly improved the performance in the third analysis. Both figures clearly show that the third analysis, the data resampling technique, outperformed the other two techniques for the decision tree and random forest classifiers.

**Table 2.** Results of three data analysis techniques.

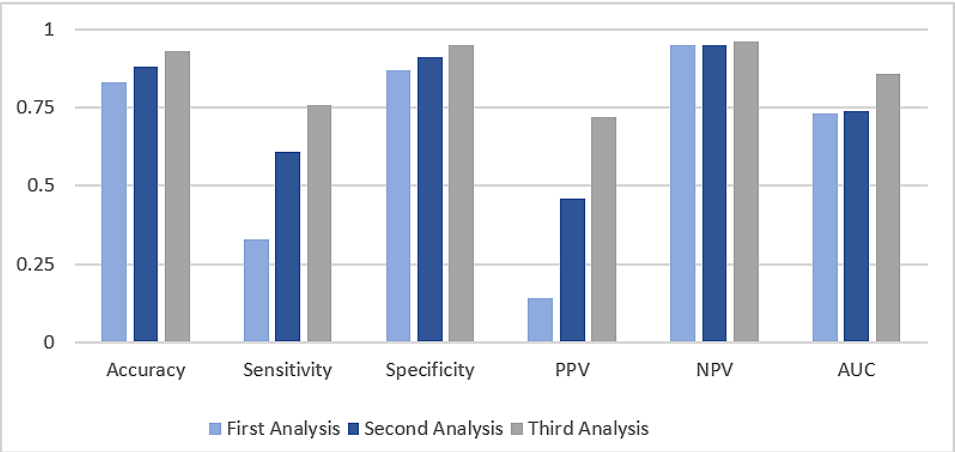
| Technique and classifier | Accuracy | Sensitivity | Specificity | PPV[a] | NPV[b] | AUC[c] |
|---|---|---|---|---|---|---|
| **Without data resampling** | | | | | | |
| Naïve Bayes | 0.82 | 0.34 | 0.88 | 0.27 | 0.91 | 0.76 |
| BayesNet | 0.82 | 0.38 | 0.89 | 0.37 | 0.90 | 0.88 |
| Decision tree | 0.83 | 0.33 | 0.87 | 0.14 | 0.95 | 0.73 |
| Random forest | 0.86 | 0.55 | 0.86 | 0.01 | 0.99 | 0.87 |
| **Data imputation** | | | | | | |
| Naïve Bayes | 0.81 | 0.32 | 0.88 | 0.25 | 0.91 | 0.74 |
| BayesNet | 0.86 | 0.53 | 0.92 | 0.54 | 0.92 | 0.85 |
| Decision tree | 0.88 | 0.61 | 0.91 | 0.46 | 0.95 | 0.74 |
| Random forest | 0.90 | 0.89 | 0.90 | 0.33 | 0.99 | 0.85 |
| **Data resampling** | | | | | | |
| Naïve Bayes | 0.82 | 0.33 | 0.88 | 0.29 | 0.90 | 0.74 |
| BayesNet | 0.87 | 0.53 | 0.93 | 0.57 | 0.92 | 0.85 |
| Decision tree | 0.93 | 0.76 | 0.95 | 0.72 | 0.96 | 0.86 |
| Random forest | 0.96 | 0.97 | 0.96 | 0.75 | 0.99 | 0.97 |

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]AUC: area under the curve.

**Figure 3.** Performance comparison among three data selection techniques for the decision tree model. AUC: area under the curve; NPV: negative predictive value; PPV: positive predictive value.






XSL•FO
RenderX

**Figure 4.** Performance comparison among three data selection techniques for the random forest model. AUC: area under the curve; NPV: negative predictive value; PPV: positive predictive value.

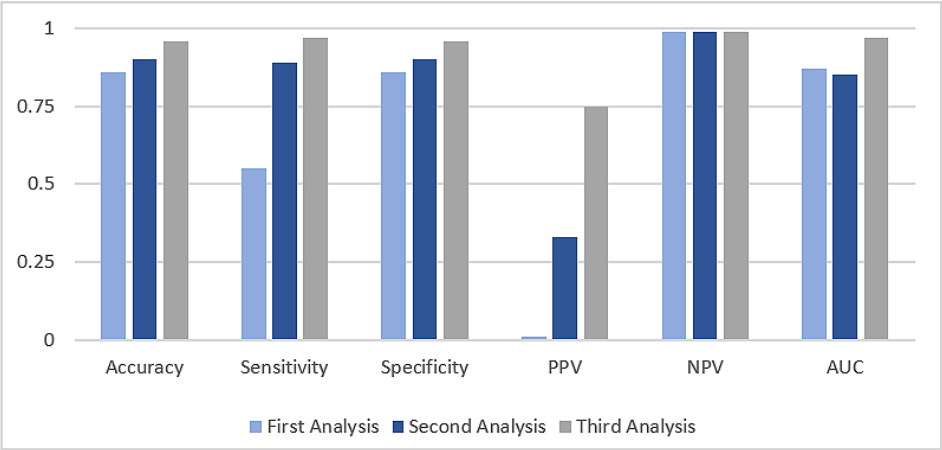



Table 3 shows the results from Pearson correlation analysis of the independent predictors.

**Table 3.** Pearson correlation coefficient values of independent predictors.

| Independent predictor of stroke | Pearson correlation coefficient (*r*) |
|---|---|
| Age | 0.26 |
| Gender | 0.13 |
| Red cell distribution width (%) | 0.18 |
| Lymphocytes (%) | 0.15 |
| Red blood cell folate (ng/mL) | 0.13 |
| Segmented neutrophils (%) | 0.12 |
| Hemoglobin (g/dL) | 0.11 |
| Red blood cell count (million cells/µL) | 0.11 |
| Hematocrit (%) | 0.09 |
| Lymphocytes (1000 cells/µL) | 0.08 |
| Segmented neutrophils (1000 cell/µL) | 0.07 |

# Discussion

## Principal Findings

From the previous section, we noticed that our models had the potential to perform stroke prediction using lab test data. Our results show that the random forest model was the best classifier after conducting the data resampling technique.

Also, several observations can be made from the results in Table 3. We identified nine lab tests, in addition to age and gender, that effectively correlated with stroke occurrence. These correlations were calculated using the Pearson correlation coefficient. These results align with other research that showed a linear relationship between some of these variables and stroke. Several studies have shown that age is correlated with the risk of stroke. According to Muntner et al [2], stroke incidence doubles after the age of 45 years, and 70% of all strokes occur over the age of 65 years. Many studies have investigated the relationship between baseline RDW and stroke. They found that elevated RDW is a risk factor in ischemic stroke [12,13,25]. One of the novel correlations that were found in this study is the lymphocyte percentage. Lymphocytes are white blood cells, including B cells, T cells, and natural killer cells. Lymphocyte percentage is positively associated with stroke occurrence. There have been no studies suggesting that lymphocyte percentage can be a predictor of stroke, but different studies have examined the use of immune cells as biomarkers to predict stroke outcome [26,27]. There is one study that showed a negative correlation between hematocrit and stroke occurrence [10]. Folate deficiency has various clinical manifestations. Our finding that serum folate level was correlated with the risk of stroke is in line with the finding of Giles et al [14], who found that a serum folate concentration of ≤9.2 nmol/L may slightly increase the risk for ischemic stroke. Other studies have shown that folic acid therapy involving folic acid, vitamin B12, and vitamin B6 reduced the risk of ischemic stroke [15,28]. Neutrophils, which are normally the most abundant circulating white blood cells and respond quickly to infection, also contribute to the main processes causing an ischemic stroke, as they facilitate the development of blood clots. Neutrophils are, therefore, also of



XSL•FO
RenderX

considerable importance as targets for treating and preventing ischemic stroke [29]. A study by Sughrue et al [10] produced results similar to ours regarding the positive association between neutrophils and stroke occurrence. Hemoglobin levels can predict the risk of stroke. Observational studies have reported an independent association between red blood cell count, hematocrit, and hemoglobin concentration and the risk of developing stroke [30,31].

The correlations between these different lab tests and stroke were found in several studies. However, this is the first study that used all of these different attributes to build a prediction model using machine learning algorithms. Our results showed that a prediction model can be created using the random forest algorithm and could achieve an accuracy of 0.96.

## Conclusions

Machine learning applications are becoming more widely used in the health care sector. The prediction of stroke using machine learning algorithms has been studied extensively. However, no previous work has explored the prediction of stroke using lab tests. The results of several laboratory tests are correlated with stroke. Building a prediction model that can predict the risk of stroke from lab test data could save lives. In this study, we created a prediction model using the random forest algorithm and achieved a 96% accuracy rate. The model can be integrated with electronic health records to provide a real-time prediction of stroke from lab tests. Because of the nature of the data, we could not predict the type of stroke: hemorrhagic or ischemic. In future studies, we aim to use data that provide information about different types of stroke to build prediction models for each type.

## Acknowledgments

EMA conducted the research design, data collection, and data analysis and wrote the original draft. AA assisted with the literature review of the lab tests. JL revised and edited the original draft and provided guidance throughout the whole research process. This study received no external funding.

## Conflicts of Interest

None declared.

## References

1. Sacco RL, Kasner SE, Broderick JP, Caplan LR, Connors JJB, Culebras A, American Heart Association Stroke Council, Council on Cardiovascular Surgery and Anesthesia, Council on Cardiovascular Radiology and Intervention, Council on Cardiovascular and Stroke Nursing, Council on Epidemiology and Prevention, Council on Peripheral Vascular Disease, Council on Nutrition, Physical Activity and Metabolism. An updated definition of stroke for the 21st century: A statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2013 Jul;44(7):2064-2089. [doi: 10.1161/STR.0b013e318296aeca] [Medline: 23652265]
2. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2019 Update: A report from the American Heart Association. Circulation 2019 Mar 05;139(10):e56-e528 [FREE Full text] [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
3. European Stroke Initiative Executive Committee, EUSI Writing Committee, Olsen TS, Langhorne P, Diener HC, Hennerici M, et al. European Stroke Initiative Recommendations for Stroke Management – Update 2003. Cerebrovasc Dis 2003;16(4):311-337 [FREE Full text] [doi: 10.1159/000072554] [Medline: 14584488]
4. Boden-Albala B, Sacco RL. Lifestyle factors and stroke risk: Exercise, alcohol, diet, obesity, smoking, drug use, and stress. Curr Atheroscler Rep 2000 Mar;2(2):160-166. [doi: 10.1007/s11883-000-0111-3] [Medline: 11122740]
5. Arnett D, Blumenthal R, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2019 Sep 10;74(10):e177-e232. [doi: 10.1016/j.jacc.2019.03.010] [Medline: 30894318]
6. Manuel DG, Tuna M, Perez R, Tanuseputro P, Hennessy D, Bennett C, et al. Predicting stroke risk based on health behaviours: Development of the Stroke Population Risk Tool (SPoRT). PLoS One 2015;10(12):e0143342 [FREE Full text] [doi: 10.1371/journal.pone.0143342] [Medline: 26637172]
7. Lee J, Lim H, Kim D, Shin S, Kim J, Yoo B, et al. The development and implementation of stroke risk prediction model in National Health Insurance Service's personal health record. Comput Methods Programs Biomed 2018 Jan;153:253-257 [FREE Full text] [doi: 10.1016/j.cmpb.2017.10.007] [Medline: 29157457]
8. Nielsen A, Hansen MB, Tietze A, Mouridsen K. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. Stroke 2018 Jun;49(6):1394-1401. [doi: 10.1161/strokeaha.117.019740]
9. Rondina JM, Filippone M, Girolami M, Ward NS. Decoding post-stroke motor function from structural brain imaging. Neuroimage Clin 2016;12:372-380 [FREE Full text] [doi: 10.1016/j.nicl.2016.07.014] [Medline: 27595065]



XSL•FO
RenderX

10. Sughrue T, Swiernik MA, Huang Y, Brody JP. Laboratory tests as short-term correlates of stroke. BMC Neurol 2016 Jul 21;16:112 [FREE Full text] [doi: 10.1186/s12883-016-0619-y] [Medline: 27439507]
11. Farah R, Samra N. Mean platelets volume and neutrophil to lymphocyte ratio as predictors of stroke. J Clin Lab Anal 2018 Jan;32(1):1-4 [FREE Full text] [doi: 10.1002/jcla.22189] [Medline: 28303662]
12. Feng G, Li H, Li Q, Fu Y, Huang R. Red blood cell distribution width and ischaemic stroke. Stroke Vasc Neurol 2017 Sep;2(3):172-175 [FREE Full text] [doi: 10.1136/svn-2017-000071] [Medline: 28989807]
13. Kaya A, Isik T, Kaya Y, Enginyurt O, Gunaydin ZY, Iscanli MD, et al. Relationship between red cell distribution width and stroke in patients with stable chronic heart failure: A propensity score matching analysis. Clin Appl Thromb Hemost 2015 Mar;21(2):160-165 [FREE Full text] [doi: 10.1177/1076029613493658] [Medline: 23804231]
14. Giles WH, Kittner SJ, Anda RF, Croft JB, Casper ML. Serum folate and risk for ischemic stroke. First National Health and Nutrition Examination Survey epidemiologic follow-up study. Stroke 1995 Jul;26(7):1166-1170. [doi: 10.1161/01.str.26.7.1166] [Medline: 7604408]
15. Qin X, Li J, Spence JD, Zhang Y, Li Y, Wang X, et al. Folic acid therapy reduces the first stroke risk associated with hypercholesterolemia among hypertensive patients. Stroke 2016 Nov;47(11):2805-2812. [doi: 10.1161/strokeaha.116.014578]
16. National Health and Nutrition Examination Survey. Centers for Disease Control and Prevention. URL: https://wwwn.cdc.gov/nchs/nhanes/Default.aspx [accessed 2020-04-10]
17. Sohan F, Rahman SSMM, Munna TA, Allayear SM, Rahman H, Rahman M. NStackSenti: Evaluation of a multi-level approach for detecting the sentiment of users. In: Proceedings of the 4th International Conference on Next Generation Computing Technologies. 2018 Presented at: 4th International Conference on Next Generation Computing Technologies; November 21-22, 2018; Dehradun, India p. 38-48. [doi: 10.1007/978-981-15-1718-1_4]
18. Marasinghe MG, Koehler KJ. Statistical Data Analysis Using SAS: Intermediate Statistical Methods. 2nd edition. Cham, Switzerland: Springer International Publishing AG; 2018.
19. Dugan TM, Mukhopadhyay S, Carroll A, Downs S. Machine learning techniques for prediction of early childhood obesity. Appl Clin Inform 2017 Dec 19;06(03):506-520. [doi: 10.4338/aci-2015-03-ra-0036]
20. Li X, Tao S, Jamal-Omidi S, Huang Y, Lhatoo SD, Zhang G, et al. Detection of postictal generalized electroencephalogram suppression: Random forest approach. JMIR Med Inform 2020 Feb 14;8(2):e17061 [FREE Full text] [doi: 10.2196/17061] [Medline: 32130173]
21. Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 4th edition. London, UK: Springer-Verlag; 2014.
22. Hsu C, Liu C, Tain Y, Kuo C, Lin Y. Machine learning model for risk prediction of community-acquired acute kidney injury hospitalization from electronic health records: Development and validation study. J Med Internet Res 2020 Aug 04;22(8):e16903 [FREE Full text] [doi: 10.2196/16903] [Medline: 32749223]
23. Du Z, Yang Y, Zheng J, Li Q, Lin D, Li Y, et al. Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: Model development and performance evaluation. JMIR Med Inform 2020 Jul 06;8(7):e17257 [FREE Full text] [doi: 10.2196/17257] [Medline: 32628616]
24. Sohan M, Kabir M, Jabiullah M, Rahman SSMM. Revisiting the class imbalance issue in software defect prediction. In: Proceedings of the 2nd International Conference on Electrical, Computer and Communication Engineering. 2019 Presented at: International Conference on Electrical, Computer and Communication Engineering; February 7-9, 2019; Cox's Bazar, Bangladesh p. 1-6. [doi: 10.1109/ecace.2019.8679382]
25. Song S, Hua C, Dornbors D, Kang R, Zhao X, Du X, et al. Baseline red blood cell distribution width as a predictor of stroke occurrence and outcome: A comprehensive meta-analysis of 31 studies. Front Neurol 2019;10:1237 [FREE Full text] [doi: 10.3389/fneur.2019.01237] [Medline: 31849813]
26. Wang Y, Liu J, Wang X, Liu Z, Li F, Chen F, et al. Frequencies of circulating B- and T-lymphocytes as indicators for stroke outcomes. Neuropsychiatr Dis Treat 2017 Oct;13:2509-2518. [doi: 10.2147/ndt.s148073]
27. Liesz A, Hu X, Kleinschnitz C, Offner H. Functional role of regulatory lymphocytes in stroke. Stroke 2015 May;46(5):1422-1430. [doi: 10.1161/strokeaha.114.008608]
28. Spence JD, Yi Q, Hankey GJ. B vitamins in stroke prevention: Time to reconsider. Lancet Neurol 2017 Sep;16(9):750-760. [doi: 10.1016/S1474-4422(17)30180-1] [Medline: 28816120]
29. Jickling GC, Liu D, Ander BP, Stamova B, Zhan X, Sharp FR. Targeting neutrophils in ischemic stroke: Translational insights from experimental studies. J Cereb Blood Flow Metab 2015 Jun;35(6):888-901 [FREE Full text] [doi: 10.1038/jcbfm.2015.45] [Medline: 25806703]
30. Kim M, Jee SH, Yun JE, Baek SJ, Lee D. Hemoglobin concentration and risk of cardiovascular disease in Korean men and women - The Korean Heart Study. J Korean Med Sci 2013 Sep;28(9):1316-1322 [FREE Full text] [doi: 10.3346/jkms.2013.28.9.1316] [Medline: 24015036]
31. Chang Y, Hung S, Ling W, Lin H, Li H, Chung S. Association between ischemic stroke and iron-deficiency anemia: A population-based study. PLoS ONE 2013 Dec 9;8(12):e82952. [doi: 10.1371/journal.pone.0082952]




XSL•FO
RenderX

## Abbreviations

**AUC:** area under the curve
**CDC:** Centers for Disease Control and Prevention
**MPV:** mean platelet volume
**NCHS:** National Center for Health Statistics
**NHANES:** National Health and Nutrition Examination Survey
**NLR:** neutrophil-to-lymphocyte ratio
**NPV:** negative predictive value
**PPV:** positive predictive value
**RDW:** red cell distribution width
**ROC:** receiver operating characteristic
**SEQN:** sequence number
**WEKA:** Waikato Environment for Knowledge Analysis

*Edited by G Eysenbach; submitted 12.08.20; peer-reviewed by T Taveira-Gomes, K Kades, A Sule; comments to author 15.09.20; revised version received 14.11.20; accepted 15.10.21; published 02.12.21.*

*Please cite as:*
*Alanazi EM, Abdou A, Luo J*
*Predicting Risk of Stroke From Lab Tests Using Machine Learning Algorithms: Development and Evaluation of Prediction Models*
*JMIR Form Res 2021;5(12):e23440*
*URL: https://formative.jmir.org/2021/12/e23440*
*doi:10.2196/23440*
*PMID:34860663*

©Eman M Alanazi, Aalaa Abdou, Jake Luo. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Using Artificial Neural Network Condensation to Facilitate Adaptation of Machine Learning in Medical Settings by Reducing Computational Burden: Model Design and Evaluation Study

Dianbo Liu[1,2], PhD; Ming Zheng[2], BSc; Nestor Andres Sepulveda[2], PhD

[1]Department of Biomedical Informatics, Harvard University, Boston, MA, United States

[2]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States

**Corresponding Author:**
Dianbo Liu, PhD
Department of Biomedical Informatics
Harvard University
1 Autumn Street
Boston, MA, 02215
United States
Phone: 1 6177101859
Email: dianbo@mit.edu

## Abstract

**Background:** Machine learning applications in the health care domain can have a great impact on people's lives. At the same time, medical data is usually big, requiring a significant number of computational resources. Although this might not be a problem for the wide adoption of machine learning tools in high-income countries, the availability of computational resources can be limited in low-income countries and on mobile devices. This can limit many people from benefiting from the advancement in machine learning applications in the field of health care.

**Objective:** In this study, we explore three methods to increase the computational efficiency and reduce model sizes of either recurrent neural networks (RNNs) or feedforward deep neural networks (DNNs) without compromising their accuracy.

**Methods:** We used inpatient mortality prediction as our case analysis upon review of an intensive care unit dataset. We reduced the size of RNN and DNN by applying pruning of "unused" neurons. Additionally, we modified the RNN structure by adding a hidden layer to the RNN cell but reducing the total number of recurrent layers to accomplish a reduction of the total parameters used in the network. Finally, we implemented quantization on DNN by forcing the weights to be 8 bits instead of 32 bits.

**Results:** We found that all methods increased implementation efficiency, including training speed, memory size, and inference speed, without reducing the accuracy of mortality prediction.

**Conclusions:** Our findings suggest that neural network condensation allows for the implementation of sophisticated neural network algorithms on devices with lower computational resources.

***(JMIR Form Res 2021;5(12):e20767)*** doi:10.2196/20767

**KEYWORDS**

artificial neural network; electronic medical records; parameter pruning; machine learning; computational burden;

## Introduction

Machine learning applications for health care can have a great impact on people's lives. Currently, the possibilities for machine learning in health care include diagnostic systems, biochemical analysis, image analysis, and drug development. One of the most significant challenges in using machine learning for health care applications is that data is usually huge and sparse, requiring important computational resources, especially for overparameterized deep neural networks (DNNs). Consequently, the availability of computational resources to use such tools can limit their widespread use, such as for people who live in low-income countries and for those who want to run diagnostic apps on their own mobile devices.

In this study, we set in-hospital mortality prediction as a case study to explore the various ways of improving efficiency (ie, training speed, memory size, and inference speed) of neural network–based algorithms. Mortality prediction is a well-tried



XSL•FO
RenderX

medical machine learning application wherein the mortality of a patient after being transferred to the intensive care unit (ICU) can be predicted based on their vital signs, laboratory tests, demographics, and other factors. Mortality prediction is important in clinical settings because such a prediction can help determine the declining state and need for intervention. We built baseline models with either recurrent neural network (RNN) or dense neural network architectures, based on which we explored efficiency improvements via neural network condensation without sacrificing the prediction accuracy. An RNN is a class of artificial neural networks wherein connections between nodes form a directed graph along a temporal sequence that consider a sequence of input in a recurrent manner. RNNs are widely used in clinical informatics in tasks such as temporal data analysis and clinical natural language processing.

Reduction of complexity and improvement of efficiency of artificial neural networks is an active field of research, wherein a wide range of methods have been explored. One representative example is neural network pruning, wherein a fraction of weights is removed from the trained model and the "lottery ticket" is found when the remaining weight can still be quickly trained with competitive loss and accuracy [1-3]. There are more fancy pruning approaches where the authors use another neural network to learn and conduct the best pruning decisions considering the network to be pruned (ie, the backbone neural network). For example, Lin et al [4] developed a method called runtime neural pruning to model their pruning process as a Markov decision process and use reinforcement learning for training via an additional RNN. Zhong et al [5], on the other hand, used long short-term memory (LSTM) to guide an end-to-end pruning of the backbone neural network. Some other previous works have converted the neural network condensation into an optimization problem where parameters are penalized under some norm [6-9].One RNN-specific condensation method is that instead of embedding information into multiple recurrent layers, we only use one recurrent layer but extend the capacity of the RNN unit (cell) by incorporating more hidden layers within the cell. Dai et al [10] showed that DNNs were inserted between the recurrent layer and the input (masking) layer for each gate in the LSTM to form an LSTM embedded with hidden layers (ie, hLSTM). Such an architecture can, in principle, be more efficient (ie, fewer number of parameters and higher training speed). There is another posttraining condensation method called *quantization*, wherein parameters originally stored in a 32-bit floating point format are forcibly converted to 8 fixed bits [11]. Other methods used for neural network condensation include, but are not limited to, binarization of neural networks [12], knowledge distillation [13], and Huffman coding [14]. In this paper, we describe the use of hLSTM, neural network pruning, and quantization to condense the size of neural networks and increase speed while maintaining their prediction accuracy.

## Methods

### Intensive Care Unit Data

We used the Medical Information Mart for Intensive Care-III (MIMIC-III) critical care database for the implementation of our models [15]. In all, 53,423 distinct hospital adult patients admitted to critical care units between 2001 and 2012 are included in this database. We excluded all neonatal and pediatric patients (aged 18 years or younger at the time of ICU stay) because the physiology of pediatric critical care patients differs significantly from that of adults [16]. We also excluded any hospital admissions with multiple ICU stays or transfers between different ICU units. The final cohort comprised 33,798 unique patients, with a total of 42,276 hospital admissions and ICU stays. Of these 33,798 patients, we defined a test set of 5070 (15%) patient stays. In-hospital mortality was determined by comparing patient date of death with hospital admission and discharge times. The mortality rate within the cohort was 10.9%. The median age of adult patients was 65.8 (SD 11.3) years, and 55.9% (18,893/33,798) patients were male. A mean of 4579 (SD 721.7) charted observations and 380 (SD 215.8) laboratory measurements, as well as other static information, are available for each hospital admission.

**Table 1.** Summary of patient data (N=33,798).

| Variable | Value |
|---|---|
| Mortality during ICU[a] stay, n (%) | 3717 (10.9) |
| Age in years, median (SD) | 65.8 (11.3) |
| Male participants | 18,893 (55.9) |

[a]ICU: intensive care unit.

### Data Prepossessing

Data were collected from the MIMIC-III database. Only data from the first 48 hours were used as inputs in our analysis. For the purpose of this study, 76 features were selected for analysis (see examples listed in Textbox 1). Some features may appear multiple times (in different means or conditions) and are thus regarded as *independent* features. We resampled the time series into regularly spaced intervals. If there were multiple measurements of the same variable in the same interval, we used the value of the last measurement. We imputed the missing values using the previous value, if it exists, or a prespecified "normal" [16] value, otherwise. In addition, we added a binary mask input for each variable, indicating the time steps that contain a true (vs imputed) measurement [17]. Categorical variables were encoded using a one-hot vector at each time step. Then, the inputs were normalized by subtracting the mean and dividing it by the SD value. Statistics were calculated per variable after imputation of missing values.



XSL•FO
RenderX

**Textbox 1.** Examples of the 76 features selected for the analysis.

- pH
- Fraction-inspired oxygen
- Systolic blood pressure
- Height
- Weight
- Oxygen saturation
- Diastolic blood pressure
- Glucose
- Temperature
- Mean blood pressure
- Capillary refill rate
- Respiratory rate
- Heart rate
- Fraction inspired oxygen
- Glasgow Coma Scale–50

## Performance Metrics

Classification accuracy of all models were measured using area under the receiver operating curve AUROC (also called AUCROC) on the test set. Sizes of model were measured by the number of parameters and sizes of the saved model file. Inference speed was calculated based on time taken to make predictions on test data and was normalized per patient. We used Python 3.6, Keras 2.2.4 with TensorFlow 1.1.2, as the backend for the analysis.

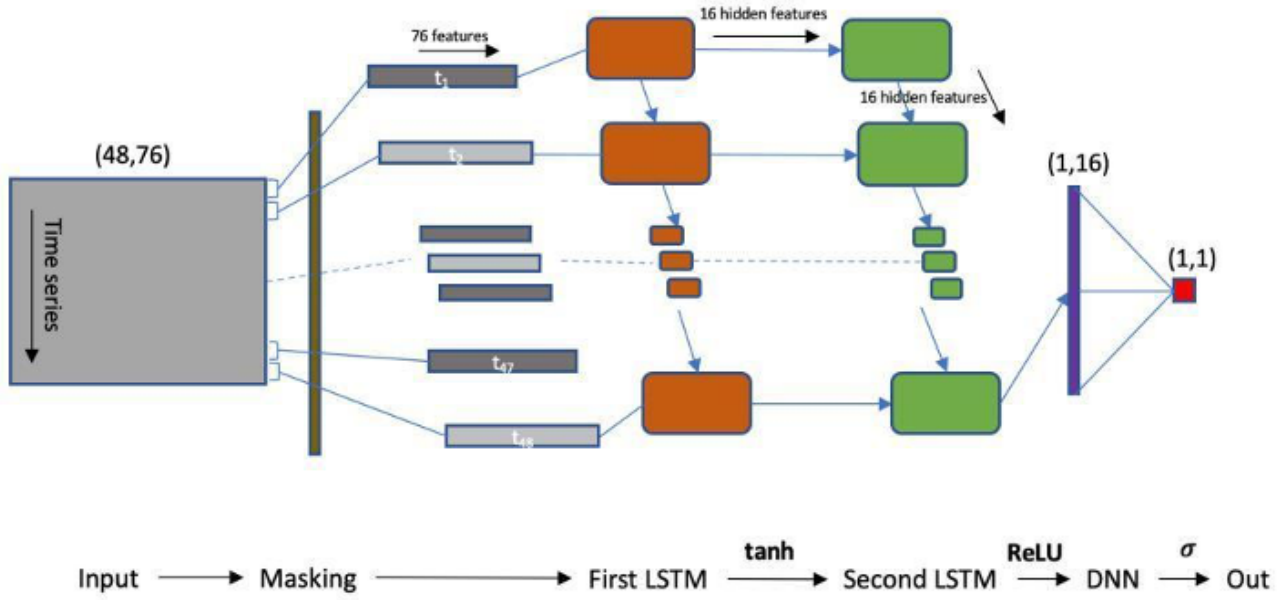
## RNN Model

Our RNN baseline model is designed as an RNN consisting of a masking layer, two LSTM layers, a dropout layer, and a dense output layer, as shown in Figure 1. We chose two layers of LSTM because, based on a literature review, we identified this structure to be the one with the best performance in the MIMIC-III mortality prediction work [16]. The masking layer masks (skips) the time step for all downstream layers if the values of input tensor at the time step are all equal to zero, which represents missing data for that time step. The first layer of LSTM takes in the original 76 features and generates a 16-feature hidden state based on the hidden state of the previous step and the new incoming observation. Then, such a hidden state is forward to the entrance of the second LSTM layer, which produces another 16-feature hidden state at each step. A dropout layer is followed by the last-step hidden state of the second LSTM layer to prevent complex coadaptations of the neurons. Finally, a dense layer is used to generate a soft 0/1 mortality prediction. The training was conducted using Adam algorithm with a dropout rate of 0.3 between layers and a learn rate of 0.001. In this study, hyperparameters were chosen by grid searching based on performance on the validation set.



XSL•FO
RenderX

**Figure 1.** Architecture of recurrent neural network baseline model. DNN: deep neural network; LSTM: long short-term memory; ReLU: rectified linear unit.




## hLSTM Model

Besides pruning upon RNN, we also tried another way by inserting an additional hidden dense layer into the inner gates of LSTM, which we called hLSTM, to improve the "power" of the LSTM. For a traditional LSTM, the inner structure is as follows:

where * is the matrix product; ⊗ is the element-wise product; $W$ represents recurrent kernel matrices of the gates; and $b$ represents corresponding bias terms. Moreover, $f$, $i$, $o$, $c$, $x$, $h$ and $c$ represent the forget gate, input gate, output gate, vector for cell updates, input, hidden state, and cell state, respectively. Subscript $t$ indicates the time step. For hLSTM, the recurrent layer in equation 1 is modified as follows:

## Feedforward DNN Model

Our baseline feed forward artificial neural network—commonly called DNN—used in this project consists of three fully connected layers, a dropout layer, and an output layer. The fully connected layers have 256, 128, and 64 neurons, respectively, and they use rectified linear unit (ReLU) as the activation function. The dropout layer has a probabilistic dropout rate of 0.5. Sigmoid function was used as activation at the output layer. The loss function used was binary cross-entropy, and the optimization algorithm used was Adam. The baseline DNN model and the pruned DNN model (pDNN) were all trained for 20 epochs, using a batch size of 8. The input into the DNN model has the same feature set as LSTM model but does not consider time series information. The values were calculated by averaging nonmissing values across time steps.

## Neural Network Pruning

All neural network prunings were conducted at the channel level, which means a neuron and all its inputs and outputs were removed from the model if the neuron is pruned. Keras surgeon library in python was used for pruning. In each layer, neurons were pruned if their mean weight across all inputs from the previous layer were below the set quantiles (ie, 25% and 50% in this study). The original model was trained for 1 epoch before pruning and was trained for another 19 epochs after pruning.

## Neural Network Quantization

Quantization was applied on the DNN model post training. Parameters, including weights and activation, originally stored in a 32-bit floating point format were converted to 8 bits using TensorFlow Lite. A uniform quantization strategy was used, as previously described [11]. Considering the range of float point values in the model to be ($F_{min}$; $F_{max}$), all the floating-point values were quantized into the range (0; 255) as 8 bits in a uniform manner, where $F_{min}$ corresponds to 0 and $F_{max}$ corresponds to 255.

The quantization process is

where $x$ is the floating-point variable, $x_q$ is the quantized variable, and

# Results

## Recurrent Artificial Neural Network Condensation: hLSTM and Pruned LSTM

Recurrent artificial neural networks (or simply, RNNs) are a group of machine learning models widely used in clinical settings that take sequential or time series information as the



XSL•FO
RenderX

input. However, training of RNNs and running inference from RNNs are relatively computationally intensive. In order to enable the machine learning algorithms to be used on devices with limited computational power, such as those in high-income countries and on mobile devices, we used three strategies to reduce the storage size of the model and to increase the speed of training and inference (Figure 2).

We built a baseline RNN using two layers of LSTM neurons to predict ICU mortality rates using MIMIC-III dataset [15]. After training, the baseline RNN model achieved a decent performance of AUROC of 0.85 (Table 2).

**Figure 2.** Neural network condensation methods. (A) Hidden-layer long short-term memory (LSTM). Instead of single fix layer nonlinearity for gate control of LSTM, multiple layer neural network with ReLu as activation were used to enhance the gate controls. In this way, fewer layers of LSTMs were needed to build a model with similar performance. (B) A large portion of parameters in artificial neural networks are redundant. We pruned 50% of the channels (neurons) with the lowest weights in each layer to reduce size and complexity of the neural network. (C) Most artificial neuron network implementation in research settings uses 32- or 64-bit floating points for model parameters. We quantized the parameters to 8 bits after training to reduce sizes of the models. DNN: dense neural network.

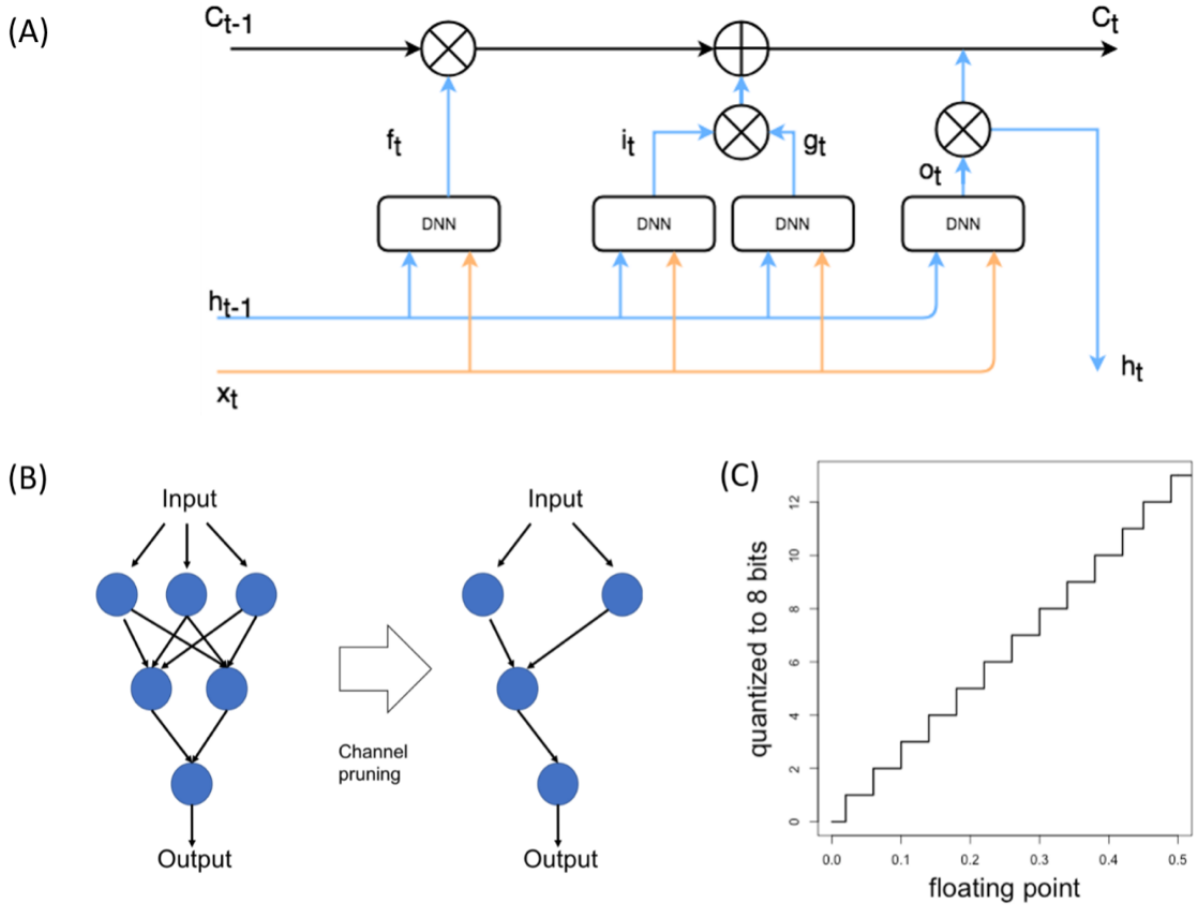



**Table 2.** Recurrent neural network condensation.

| Model | Parameters, n | File size (kb) | Inference (seconds per sample) | Training time (seconds; 20 epochs) | Test AUROC[a] (last epoch) |
|---|---|---|---|---|---|
| Baseline LSTM[b] | 8081 | 129 | 523 | 4890 | 0.836 |
| Pruned LSTM | 3273 | 73 | 318 | 4990 | 0.853 |
| Hidden-layer LSTM | 6993 | 111 | 254 | 3000 | 0.860 |

[a]AUROC: area under the receiver operating curve.

[b]LSTM: long short-term memory.

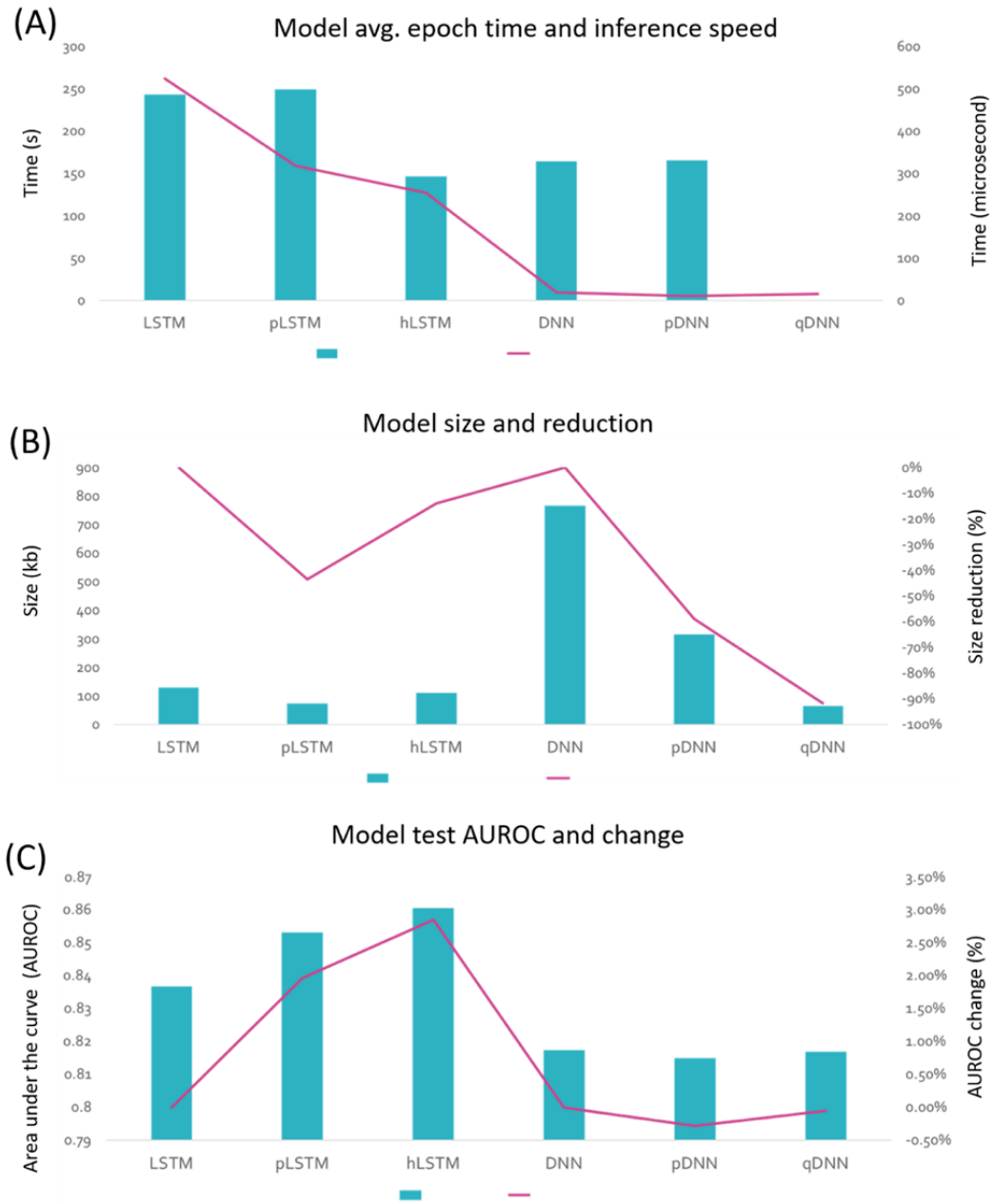
The first strategy was to modify the LSTM cell to increase the representation power of each layer. We modified the original neural network structure and added an additional hidden layer into the original LSTM class, wherein one additional layer called "hidden kernel" was inserted between the input kernel and the recurrent kernel (see equation 4). By using this strategy, we replaced the old 2-layer LSTM with only one layer of hLSTM, such that we simplified the overall structure by trying to embed the same quantity of information in this single "condensate" layer.



XSL•FO
RenderX

Both the baseline model and the hLSTM model with only one layer of hLSTM are trained under the same settings. The comparison of AUROC and accuracy is shown in Figure 3. The number of parameters for these two models are listed in Table 2. This simplified model with a single layer of hLSTM beats the baseline model 2-fold in training speed, achieving a 32% reduction in parameter numbers while simultaneously maintaining a higher AUROC at the same time.

**Figure 3.** Accuracy, model size, and inference speed of feedforward recurrent and neural networks (RNNs) after different types of condensation. (A) Area under the receiver operating characteristic curve (AUROC) of various models. (B) Various model sizes in memory. (C) Inference speed of various models. Models included the RNN baseline model with two layers of long-term short memory (LSTM), pruned LSTM (pLSTM) model, and one hidden layer inserted in LSTM (hLSTM); deep neural network (DNN) baseline model; pruned DNN (pDNN) model; quantized DNN (qDNN) model.

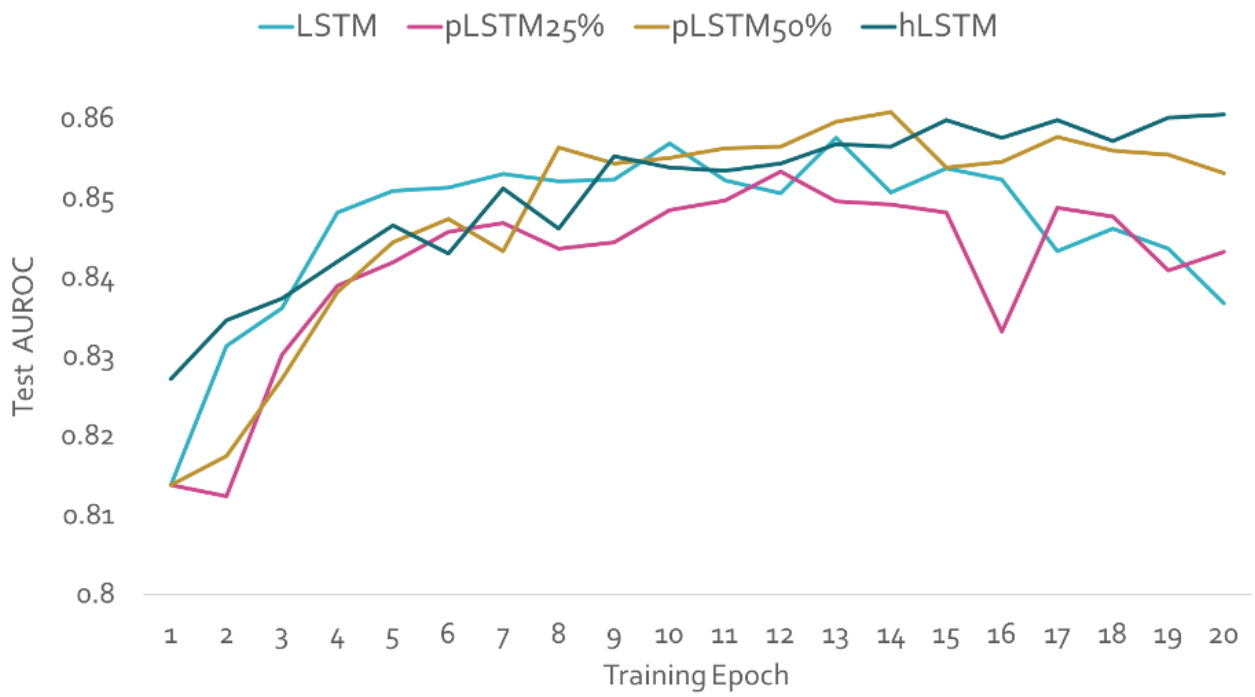



Another method to condense RNN models is pruning, in which some unessential neurons of the RNN model are removed to minimize model size and increase speed. About 50% of LSTM neurons with lowest weights in each hidden layer were pruned after the first epoch of training. The pruned LSTM only has half of the number parameters of the original LSTM, but it achieves a similar level of accuracy, yielding an AUROC of 0.85 (Figure 4). The inference speed of pruned LSTM also doubled compared with the original LSTM (Table 2).



XSL•FO
RenderX

**Figure 4.** Test area under the receiver operating curve (AUROC) by training epoch for recurrent neural network (RNN) models. Evolution of different RNN models over training epochs on test data. The percentage next to the pruned long short-term memory (pLSTM) model indicates the pruned percentile. hLSTM: hidden-layer long short-term memory; LSTM: long short-term memory.




### Feedforward Neural Network Condensation: Pruning and Quantization

Feedforward neural network, or commonly called DNN if it has multiple hidden layers, is another widely used form of machine learning in clinical settings. We trained DNN with 3 hidden layers, consisting of 256, 128, and 64 neurons in each layer, to enable ICU mortality prediction. The baseline DNN achieved an AUROC of 0.82, using patient data collected within the first 48 hours after admission. We explored two methods to condense the size of the DNN. The first method, called pruning, used the pruning strategy as in RNN; for this purpose, 50% of the channels were pruned after the first epoch of training, the prediction accuracy of the pDNN maintained at the same level as the original DNN, and the inference speed doubled (Table 3). The second strategy involved quantization, which refers to the process of reducing the number of bits that represent a number. In the context of this project, the predominant numerical format used was a 32-bit floating point. We used an after-training-quantization strategy to represent the parameters of the DNN model using 8-bit integers (ie, quantized DNN or qDNN). This method reduced storage size of the DNN model by 5 times without incurring significant loss in accuracy (Table 3). We also compared the overall performances of DNN condensation with those of RNN, as shown in Figure 3.

**Table 3.** Feedforward neural network condensation.

| Model | Parameters, n | File size (kb) | Inference (seconds per sample) | Training time (seconds;20 epochs) | Test AUROC[a] (last epoch) |
|---|---|---|---|---|---|
| Baseline DNN[b] | 60,929 | 767 | 20 | 3300 | 0.82 |
| Pruned DNN | 27,312 | 315 | 10 | 3310 | 0.81 |
| Quantized DNN | 60,929 | 64 | 15 | N/A[c] | 0.82 |

[a]AUROC: area under the receiver operating curve.

[b]DNN:

[c]N/A: not applicable.

## Discussion

In this study, we were able to use data from the MIMIC-III database [15] to train in-hospital mortality neural network models with high accuracy and conduct model condensation with different methods to gain efficiency (eg, memory size reduction and increased speed) without compromising accuracy. We implemented different neural network architectures for both RNNs and dense neural networks; thus, our methods can add value in both settings. We pioneered RNN pruning with clinical implementation and our condensation treatments aiming at higher efficiency can be extended to other medical applications using similar data, and probably to nonmedical applications as well. In addition, in medical settings, model calibration is conducted after initial model training. Calibration can be



XSL•FO
RenderX

conducted using various training schemes and early stopping strategies. The model condensation method proposed in this study significantly reduces the number of parameters and will help make model calibration easier. The major limitation of the neural network condensation method is that although our proposed method significantly reduces the sizes of different models and their computational costs in training, the final model sizes after condensation are still proportional to the original model sizes. Therefore, if further model size reduction is warranted, a combination of better model design and neural network condensation will be required.

## Acknowledgments

We thank the authors of the published article "Multitask Learning and Benchmarking with Clinical Time Series Data" for providing the basic code for data preprocessing from Medical Information Mart for Intensive Care-III via GitHub. We thank Professor David Sontag (Massachusetts Institute of Technology, Cambridge, MA) for their helpful discussion.

## Conflicts of Interest

None declared.

## References

1. Frankle J, Carbin M. The lottery ticket hypothesis: finding sparse, trainable neural networks. arXiv. Preprint posted online on March 9, 2018. [FREE Full text]
2. See A, Luong MT, Manning CD. Compression of neural machine translation models via pruning. : Association for Computational Linguistics; 2016 Presented at: The 20th SIGNLL Conference on Computational Natural Language Learning; August 2016; Berlin, Germany p. 291-301 URL: https://aclanthology.org/K16-1029/
3. Babaeizadeh M, Smaragdis P, Campbell RH. NoiseOut: a simple way to prune neural networks. arXiv. Preprint posted online on Nov 18, 2016. [FREE Full text]
4. Lin J, Rao Y, Lu J, Zhou J. Runtime neural pruning. In: Advances in Neural Information Processing Systems 30. 2017 Presented at: NIPS 2017; December 4-9, 2017; Long Beach p. 2181-2191 URL: https://papers.nips.cc/paper/2017/hash/a51fb975227d6640e4fe47854476d133-Abstract.html
5. Zhong J, Ding G, Guo Y, Han J, Wang B. Where to prune: using LSTM to guide end-to-end pruning. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. 2018 Presented at: International Joint Conferences on Artificial Intelligence Organization; July 13-19, 2018; Stockholm p. 3205-3211 URL: https://www.ijcai.org/proceedings/2018/0445.pdf [doi: 10.24963/ijcai.2018/445]
6. Carreira-Perpinn M, Idelbayev Y. Learning-compression algorithms for neural net pruning. 2018 Presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT p. 8532-8541. [doi: 10.1109/cvpr.2018.00890]
7. Guo Y, Yao A, Chen Y. Dynamic network surgery for efficient DNNs. In: Advances in Neural Information Processing Systems 30. 2016 Presented at: NIPS 2017; December 4-9, 2017; Long Beach. p. 1379-1387.
8. Aghasi A, Abdi A, Nguyen N, Romberg J. Net-trim: Convex pruning of deep neural networks with performance guarantee. In: Advances in Neural Information Processing Systems 30. 2017 Presented at: NIPS 2017; December 4-9, 2017; Long Beach p. 3177-3186.
9. Louizos C, Welling M, Kingma D. Learning Sparse Neural Networks through L0 Regularization. arXiv. Preprint posted online on December 4, 2017. [FREE Full text]
10. Dai X, Yin H, Jha NK. Grow and prune compact, fast, and accurate LSTMs. IEEE Trans Comput 2020 Mar 1;69(3):441-452. [doi: 10.1109/tc.2019.2954495]
11. Krishnamoorthi R. Quantizing deep convolutional networks for efficient inference: A whitepaper. arXiv. Preprint posted online on June 21, 2018. [FREE Full text]
12. Courbariaux M, Hubara I, Soudry D, El-Yaniv R, Bengio Y. Binarized neural networks: Training deep neural networks with weights and activations constrained to +1 or -1. arXiv. Preprint posted online on February 9, 2016. [FREE Full text]
13. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv. Preprint posted online on March 9, 2015. [FREE Full text]
14. Han S, Mao H, Dally WJ. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding. arXiv. Preprint posted online on October 1, 2015. [FREE Full text]
15. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
16. Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Sci Data 2019 Jun 17;6(1):96 [FREE Full text] [doi: 10.1038/s41597-019-0103-9] [Medline: 31209213]
17. Lipton ZC, Kale DC, Wetzel R. Modeling missing data in clinical time series with RNNs. arXiv. Preprint posted online on June 13, 2016. [FREE Full text]



XSL•FO
RenderX

## Abbreviations

**AUROC:** area under the receiver operating curve
**DNN:** deep neural network
**hLSTM:** hidden-layer long short-term memory
**ICU:** intensive care unit
**LSTM:** long short-term memory
**MIMIC-III:** Medical Information Mart for Intensive Care-III
**pDNN:** pruned deep neural network
**qDNN:** quantized deep neural network
**ReLU:** rectified linear unit
**RNN:** recurrent neural network

*Edited by G Eysenbach; submitted 28.05.20; peer-reviewed by H Deng, J Farzi, J Li; comments to author 14.07.20; revised version received 30.11.20; accepted 07.12.20; published 08.12.21.*

*Please cite as:*
*Liu D, Zheng M, Sepulveda NA*
*Using Artificial Neural Network Condensation to Facilitate Adaptation of Machine Learning in Medical Settings by Reducing Computational Burden: Model Design and Evaluation Study*
*JMIR Form Res 2021;5(12):e20767*
*URL: https://formative.jmir.org/2021/12/e20767*
*doi:10.2196/20767*
*PMID:34889747*

©Dianbo Liu, Ming Zheng, Nestor Andres Sepulveda. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Short Paper

# Campus Smoking Policies and Smoking-Related Twitter Posts Originating From California Public Universities: Retrospective Study

Joshua S Yang[1], MPH, PhD; Raphael E Cuomo[2,3], MPH, PhD; Vidya Purushothaman[2,3], MBBS, MAS, MPH; Matthew Nali[2,3,4], BA; Neal Shah[2], BS; Cortni Bardier[2,4,5], MA; Nick Obradovich[6], PhD; Tim Mackey[2,4,5], MAS, PhD

[1]Department of Public Health, California State University Fullerton, Fullerton, CA, United States
[2]Global Health Policy and Data Institute, San Diego, CA, United States
[3]Department of Anesthesiology, University of California San Diego, San Diego, CA, United States
[4]S-3 Research, San Diego, CA, United States
[5]Global Health Program, Department of Anthropology, University of California, San Diego, La Jolla, CA, United States
[6]Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany

**Corresponding Author:**
Tim Mackey, MAS, PhD
Global Health Program
Department of Anthropology
University of California, San Diego
9500 Gilman Drive
Mail Code 0505
La Jolla, CA, 92093
United States
Phone: 1 9514914161
Email: tkmackey@ucsd.edu

## Abstract

**Background:** The number of colleges and universities with smoke- or tobacco-free campus policies has been increasing. The effects of campus smoking policies on overall sentiment, particularly among young adult populations, are more difficult to assess owing to the changing tobacco and e-cigarette product landscape and differential attitudes toward policy implementation and enforcement.

**Objective:** The goal of the study was to retrospectively assess the campus climate toward tobacco use by comparing tweets from California universities with and those without smoke- or tobacco-free campus policies.

**Methods:** Geolocated Twitter posts from 2015 were collected using the Twitter public application programming interface in combination with cloud computing services on Amazon Web Services. Posts were filtered for tobacco products and behavior-related keywords. A total of 42,877,339 posts were collected from 2015, with 2837 originating from a University of California or California State University system campus, and 758 of these manually verified as being about smoking. Chi-square tests were conducted to determine if there were significant differences in tweet user sentiments between campuses that were smoke- or tobacco-free (all University of California campuses and California State University, Fullerton) compared to those that were not. A separate content analysis of tweets included in chi-square tests was conducted to identify major themes by campus smoking policy status.

**Results:** The percentage of positive sentiment tweets toward tobacco use was higher on campuses without a smoke- or tobacco-free campus policy than on campuses with a smoke- or tobacco-free campus policy (76.7% vs 66.4%, $P$=.03). Higher positive sentiment on campuses without a smoke- or tobacco-free campus policy may have been driven by general comments about one's own smoking behavior and comments about smoking as a general behavior. Positive sentiment tweets originating from campuses without a smoke- or tobacco-free policy had greater variation in tweet type, which may have also contributed to differences in sentiment among universities.

**Conclusions:** Our study introduces preliminary data suggesting that campus smoke- and tobacco-free policies are associated with a reduction in positive sentiment toward smoking. However, continued expressions and intentions to smoke and reports of



XSL•FO
RenderX

one's own smoking among Twitter users suggest a need for more research to better understand the dynamics between implementation of smoke- and tobacco-free policies and resulting tobacco behavioral sentiment.

***(JMIR Form Res 2021;5(12):e33331)*** doi:10.2196/33331

**KEYWORDS**

tobacco-free policies; social media; colleges and universities; smoking; smoking; smoking policy; campus policy; tobacco use; Twitter analysis; smoke-free; tobacco-free; Twitter; college students; students; campus; health policy

## Introduction

The number of colleges and universities with smoke- or tobacco-free campus policies has been increasing [1-4]. As of July 2020, there were an estimated 2542 completely smoke-free campus sites (including 2104 completely tobacco-free sites), 2176 of which prohibit e-cigarette use everywhere [5]. Existing evidence suggests that smoke- and tobacco-free campus policies are well-received by the campus community [6,7] and norms shift to greater disapproval of tobacco use on campus [7]. Smoking rates appear to decline after the implementation of smoke- and tobacco-free campus policies [8,9], though e-cigarette use may increase after smoking restrictions are implemented [9,10]. Comparison of policies across universities suggests that stronger policies are associated with reduced second-hand smoke exposure [11], smoking behavior [11,12], and seeing others smoking [12].

The effects of campus smoking policies on overall sentiment, particularly among young adults, are more difficult to assess given the changing tobacco and e-cigarette product landscape and differential attitudes toward policy implementation and enforcement. Geolocating social media posts to specific spatiotemporal areas allows for comparison of in situ smoking-related attitudes and behaviors, including between campuses with and without smoke- or tobacco-free campuses. Hence, geolocation information available from publicly available social media Twitter posts represents an opportunity to garner infoveillance-generated insights retrospective to the policy implementation periods.

Public 4-year universities in California provide a unique comparison for assessing tobacco-related attitudes and behaviors during policy implementation. The state has two public 4-year university systems, the University of California (UC) and California State University (CSU). As of January 2014, a systemwide tobacco-free policy went into effect on all UC campuses prohibiting tobacco use, including e-cigarettes, on campus grounds [13]. The CSU followed suit with a systemwide policy effective September 2017 [14]. The intervening period provided an opportunity to assess the impact of the different system policies on campus tobacco-related attitudes and behaviors. Hence, the goal of the study was to retrospectively compare campus climate toward tobacco use (without respect to type of smoking product) between universities with and those without smoke- or tobacco-free campus policies using geolocated data from Twitter.

## Methods

### Data Collection

Geolocated Twitter posts from 2015, a year after which the UC system had enacted its systemwide tobacco-free policy and prior to enactment in the CSU system, were used to conduct a retrospective analysis of smoking-themed tweets geofenced from UC and CSU campuses (one CSU campus—Fullerton—enacted its own smokefree policy effective August 2013). Data collection was conducted using the Twitter public application programming interface in combination with cloud computing services on Amazon Web Services. Data were filtered for messages that included geospatial coordinates enabled by users. These data were collected in JSON format and stored in a relational database, with information on the date and time of the post, hyperlink to the original tweet, text of the tweet, and geospatial coordinates.

Posts were selected for further analysis if they included any of the following keywords associated with tobacco products and behavior: *bidis*, *cigarette*, *cigarettes*, *cigarillos*, *cigars*, *ciggie*, *class*, *dip*, *e-cig*, *hookah*, *huqqa*, *joint*, *JUUL*, *kreteks*, *Marlboro*, *Newport*, *njoy*, *pipe*, *roll-up*, *shag*, *smoke*, *smoking*, *snuff*, *snus*, *tobacco*, *vape*, *vaped*, *vapejuice*, *vaper*, *vapes*, *vapine*, *vapor*, *waterpipe*, *waxpen*, or *weed*. Keywords were selected and adopted on the basis of prior studies [15,16] and because they were related to college life or tobacco products according to manual searches conducted on Twitter. For the sake of parsimony, the top 2 leading brands of cigarette sold in the United States (Marlboro and Newport), which were also the most commonly mentioned brands in tweets related to college life as revealed from manual searches conducted on Twitter, were also included in the list of keywords.

Tweets that included these keywords were then further filtered to identify those physically originating (ie, geofenced) from CSU or UC campuses using base maps of these schools available from the Stanford Prevention Research Center [17], which the authors visually assessed for concurrent validity with school boundaries in satellite imagery via Google Maps. Latitude and longitude coordinates for posts containing keywords were loaded into ArcGIS Desktop (version 10.6) and the clipping function was used to omit tweets outside of CSU or UC campus shape files.

### Quantitative Data Analysis

Human annotators manually assessed posts relating to the study theme corresponding to user-generated tobacco products or related behavior, for face validity. Human annotators were trained in tobacco research and have participated in prior research infoveillance research [18,19]. Posts were manually



XSL•FO
RenderX

annotated for positive, negative, or neutral sentiment toward smoking. These attributes were selected by 2 authors, with high interrater reliability (Cohen κ=0.96). Discrepancies were resolved through discussion among all authors. Chi-square tests were conducted to determine if there were significant differences in sentiment between campuses that were smoke- or tobacco-free (all UC campuses and CSU, Fullerton) compared to those that were not.

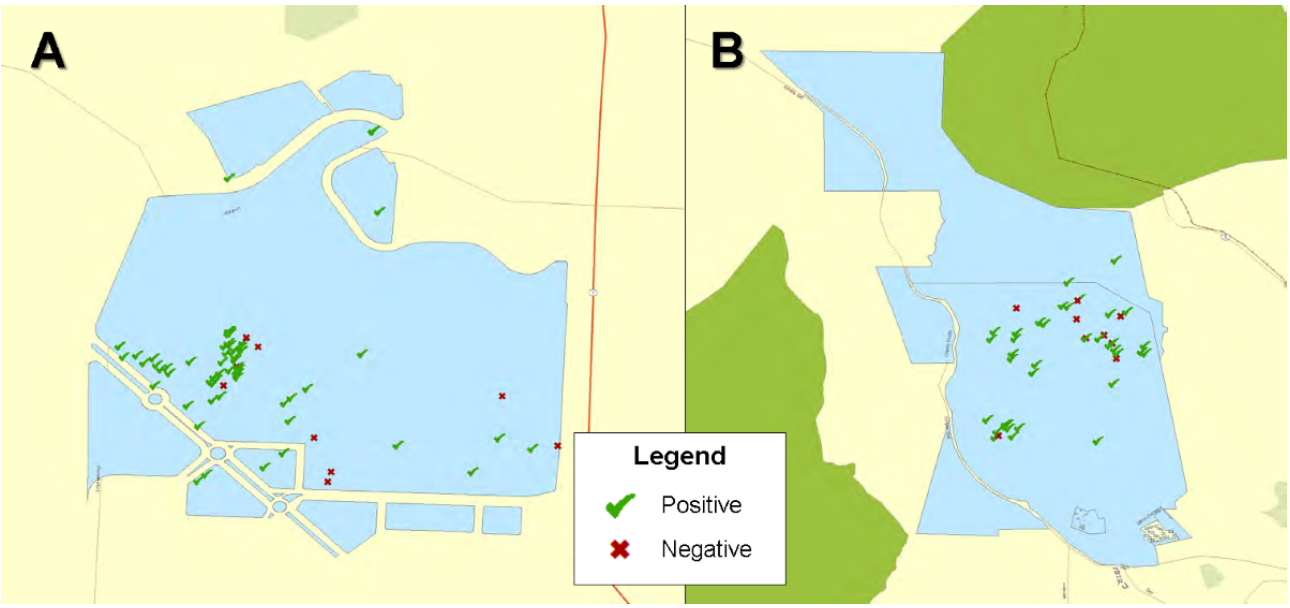
### Qualitative Content Analysis

A separate content analysis of tweets with smoking-related sentiment was conducted by the first author to identify central themes [20]. Tweets were imported into Atlas.ti (version 8) [21]. Each tweet was a recording unit and mutually exclusive. All tweets were coded for the smoking policy of the campus from which it originated (tobacco- or smoke-free vs non–tobacco- or –smoke-free). A general inductive approach was then used to develop a coding framework for tweets to assess thematic content [22]. Once the coding scheme was developed, tweets were coded and analyzed for major themes by campus smoking policy type.

## Results

A total of 42,877,339 posts with smoking-related keywords were collected from 2015, with 2837 originating from a UC or CSU system campus, and 758 of these manually verified as being associated with user-generated smoking behavior discussions. Among Twitter posts for which positive or negative sentiment was identified (396 posts from 286 unique users), 66.4% (n=89) of posts originating from smoke- and tobacco-free campuses had a positive sentiment toward tobacco products compared to 76.7% (n=201) of posts originating from campuses without smoke- or tobacco-free campus policies (*P*=.03). Figure 1 shows geocoding of positive and negative sentiment tweets from the CSU (A) and UC (B) campuses from which the greatest number of tweets originated.

**Figure 1.** Map depicting geocoding of positive and negative sentiment tweets from California State University (A) and University of California (B) campuses, which had the greatest number of tweets.




Four thematic categories of negative sentiment tweets and 9 thematic categories of positive sentiment tweets emerged from the data (Table 1). Content analysis identified 2 dominant themes among negative sentiment tweets on campuses with smoke- and tobacco-free policies: observations of others smoking on campus (likely in violation of existing policies) and displeasure associated with the smell of cigarettes or smokers (Table 2). Assessment of positive sentiment tweets between campuses suggests that the difference in positive sentiment may be driven by the larger number of general comments about one's own smoking behavior and comments about smoking as a general behavior. A greater variety in types of positive sentiment tweets originating from campuses without a smoke- or tobacco-free policy may have also contributed to differences in sentiment toward smoking between campuses with and those without smoke-free policies; these include urging others to smoke, expression of positive opinions about smoking, and attraction to people who smoke or environments where smoking occurs, though none of these alone were considered dominant themes.



XSL•FO
RenderX

**Table 1.** Thematic categories of smoking-related tweets originating from public 4-year universities by sentiment, California, 2015.

| Category | Deidentified examples |
|---|---|
| **Negative sentiment tweets** | |
| Observation of others smoking on campus | • *This [person] is smoking a hookah pen in class […] is that necessary* [a]<br>• *Thanks to all the people who smoke at [deleted]. The second hand smoke is real* [b] |
| Smell of cigarettes or smokers | • *Why did [deleted] who smells like cigarettes sit in front of me* [a]<br>• *[deleted] next to me straight smells like cigarettes & I wanna throw up* [b] |
| Support for smoke- or tobacco-free policy | • *It's a no smoking campus [deleted]. #caughingmylungsout* [a]<br>• *Want to make our campus smoke free hate [deleted]* [b] |
| Opinions against smoking | • *Smoking cigarettes is giving [deleted] a death sentence. Guhross* [b] |
| **Positive sentiment tweets** | |
| Expression of desire to smoke | • *tryna smoke but [deleted] is down* [a]<br>• *Is it too early to smoke?* [b] |
| Report of one's own smoking | • *Smoke break* [a]<br>• *[That] was [some] strong af hookah* [b] |
| Intention to smoke | • *Revising my paper […] as I wait for someone so we can go smoke* [a]<br>• *If anyone needs me I'll be smoking cigarettes […]* [b] |
| Opposition to campus smoking policy | • *[Delete] has a smoke free campus bs man* [a] |
| General comments about one's smoking behavior | • *I don't smoke anymore [delete] ... But i also don't smoke any less* [b] |
| Comments about smoking as general behavior | • *Smoke […] eat […] live […]* [b] |
| Urging others to smoke | • *@---------- go smoke with him* [b] |
| Positive opinions about smoking | • *[To] all the people smoking cigarettes at least you aren't vaping* [b] |
| Attraction to people who smoke or environments where smoking occurs | • *Oddly attracted to how [delete] look smoking a cigarette* [b] |

[a]Tweet originating from a campus with a smoke- or tobacco-free campus policy.
[b]Tweet originating from a campus without a smoke- or tobacco-free campus policy.

**Table 2.** Dominant smoking-related themes originating from public 4-year universities by sentiment and campus smoking policy type, California, 2015.

| Campus type | Negative sentiment | Positive sentiment |
|---|---|---|
| Campuses with smoke- or tobacco-free policies | • Observations of others smoking on campus (n=11)<br>• Smell of cigarettes or smokers (n=8) | • Expressions of desire to smoke (n=22)<br>• Reports of one's own smoking (n=10)<br>• Intention to smoke (n=8) |
| Campuses without smoke- or tobacco-free policies | • Opinions opposed to smoking (n=10)<br>• Observations of others smoking on campus (n=8)<br>• Smell of cigarettes or smokers (n=7) | • Expressions of desire to smoke (n=36)<br>• Reports of one's own smoking (n=23)<br>• General comments about one's own smoking behavior (n=18)<br>• Comments about smoking as a general behavior (n=18)<br>• Intention to smoke (n=14) |



XSL•FO
RenderX

## Discussion

### Principal Findings

Quantitative analysis of tweets originating from campuses of public 4-year universities in California revealed a significant difference in sentiment toward smoking by campus smoking policy status, with a higher proportion of positive sentiment tweets originating from campuses without smoke- or tobacco-free policies, but an overall high percentage of positive sentiment regardless of campus smoking policy type. Greater positive sentiment may have been driven by more tweets containing general comments about one's own smoking behavior and comments about smoking as a general behavior and a greater variety in types of positive sentiment tweets.

### Comparison With Prior Work

Negative sentiment toward smoking was identified on both campus types, but universities without a smoke- or tobacco-free policy had much higher numbers of positive sentiment tweets in two categories: general comments about one's own smoking behavior and comments about smoking as a general behavior. These differences may indicate policy passage and implementation at smoke- and tobacco-free campuses may be associated with less positive sentiment toward smoking, consistent with other studies [6,7]. Though efforts such as outreach programs and infrastructure changes may help with policy compliance [23-25], continued effort is needed to change norms around campus smoking and increase policy buy-in. These efforts should be ongoing and consistent, facilitating broad involvement of campus constituents in the policy implementation process. Additionally, twitter posts expressing a desire and intention to smoke as well as self-reported smoking suggest a need to more actively promote cessation on all campuses [24-25]. Future studies should also assess whether sentiment may significantly differ on the basis of different tobacco, e-cigarette, and other smoking product use, particularly in the context of introduction of new products and bans on products (eg, flavored products) that are popular among youth and young adults.

### Limitations

Data collected for this study were limited to Twitter users who enabled geolocation, which may have introduced bias in the volume and types of tweets or Twitter users for whom data were collected. Additionally, this study only reviewed data from a single calendar year to assess differences in sentiment between California universities with and those without smoke- or tobacco-free campus policies, though examining tweets from subsequent years (2016 and 2017) may have yielded additional user sentiment after policy implementation. Another limitation is the constrained contextual information available in some tweets (eg, such as references to "smoking" without mention of a specific product), found mainly among tweets expressing an intention to smoke, which may have referred to marijuana and not tobacco products. We also did not restrict our data set to one tweet per user, though a user expressing negative or positive sentiment may be more likely to have a similar sentiment in other tweets they post. The rationale for this approach included the detection of both second- and first-hand accounts of smoking behavior, which could relate to multiple smoking behavior instances and the possibility that this could lead to possible underestimation of expressed sentiment toward alleged violations of campus smoking bans. Finally, this is an ecological study assessing the relationship between a survey of smoking-related social media posts from college campuses and the policies on those campuses. Associations uncovered in this study may not apply to individuals, and the effect of college policies on sentiment should be further investigated in studies with a breadth of individual participants whose selection is representative of the campus population.

### Conclusions

Among Twitter users in California public universities, the overall sentiment toward tobacco products and use is high, although positive sentiment is higher on campuses without smoke- or tobacco-free campus policies. Smoke- and tobacco-free policy initiatives, including implementation, may help reduce positive sentiment toward smoking on college campuses and should be strengthened to maximally increase their impact.

### Acknowledgments

This study was partially supported by funds provided by The Regents of the University of California, Tobacco-Related Diseases Research Program (grant T29IP0465). The opinions and conclusions herein are those of the author and not necessarily represent those of The Regents of the University of California or any of its programs.

### Conflicts of Interest

None declared.

### References

1. Trad C, Bayly J, Saint-Fort L, Andrews M, Patel M, Sabado-Liwag M, et al. Adoption of tobacco- and smoke-free policies in a US national sample of postsecondary educational institutions. Am J Public Health 2018 Oct;108(10):1366-1369 [FREE Full text] [doi: 10.2105/AJPH.2018.304568] [Medline: 30138061]
2. Bayly J, Trad C, Saint-Fort L, Andrews M, Patel M, Haynie D, et al. Adoption of electronic-cigarette-free, hookah-free and American College Health Association recommended tobacco-free policies among a national sample of postsecondary



XSL•FO
RenderX

educational institutions. J Am Coll Health 2020 Jan;68(1):26-31 [FREE Full text] [doi: 10.1080/07448481.2018.1527772] [Medline: 30388931]
3. Wang TW, Tynan MA, Hallett C, Walpert L, Hopkins M, Konter D, et al. Smoke-free and tobacco-free policies in colleges and universities - United States and Territories, 2017. MMWR Morb Mortal Wkly Rep 2018 Jun 22;67(24):686-689 [FREE Full text] [doi: 10.15585/mmwr.mm6724a4] [Medline: 29927904]
4. Blake KD, Klein AL, Walpert L, Casey L, Hallett C, Douglas C, et al. Smoke-free and tobacco-free colleges and universities in the United States. Tob Control 2020 May;29(3):289-294. [doi: 10.1136/tobaccocontrol-2018-054829] [Medline: 31266902]
5. Smokefree colleges and universities continue to grow in popularity. American Nonsmokers' Rights Foundation. URL: https://no-smoke.org/at-risk-places/colleges/ [accessed 2021-12-19]
6. Ickes MJ, Rayens MK, Wiggins A, Hahn EJ. Students' beliefs about and perceived effectiveness of a tobacco-free campus policy. Policy Polit Nurs Pract 2017 Feb;18(1):17-25. [doi: 10.1177/1527154417700633] [Medline: 28558518]
7. Wray RJ, Hansen N, Ding D, Masters J. Effects of a campus-wide tobacco-free policy on tobacco attitudes, norms and behaviors among students, staff and faculty. J Am Coll Health 2021;69(8):860-871. [doi: 10.1080/07448481.2020.1711763] [Medline: 31995456]
8. Rogers CJ, Barrington-Trimis JL, Unger JB, Forster M. Changes in smoking prevalence and perception of smoking on campus before and after a smoke-free university campus policy. J Am Coll Health 2020 Jul 23:1-5. [doi: 10.1080/07448481.2020.1786097] [Medline: 32703095]
9. Leavens ELS, Lechner WV, Stevens EM, Miller MB, Meier E, Brett EI, et al. Electronic cigarette and combustible cigarette use following a campus-wide ban: prevalence of use and harm perceptions. J Am Coll Health 2020;68(4):332-335. [doi: 10.1080/07448481.2018.1551803] [Medline: 30681933]
10. Llanes K, Cabriales JA, Hernandez N, Cooper TV. Electronic cigarette use after the adoption of a tobacco-free campus policy. Addict Behav 2019 Mar;90:324-328. [doi: 10.1016/j.addbeh.2018.11.037] [Medline: 30503952]
11. Fujita S, Marteache N. Evaluation of a smoke- and tobacco-free campus policy: The issue of displacement. J Am Coll Health 2020 May 20:1-9. [doi: 10.1080/07448481.2020.1758705] [Medline: 32432986]
12. Lee JGL, Ranney LM, Goldstein AO. Cigarette butts near building entrances: what is the impact of smoke-free college campus policies? Tob Control 2013 Mar;22(2):107-112. [doi: 10.1136/tobaccocontrol-2011-050152] [Medline: 22135167]
13. Smoke and Tobacco Free Environment. University of California. URL: https://policy.ucop.edu/doc/4000371/SmokingPolicy#:~:text=The%20University%20of%20California%20prohibits,free%20effective%20January%202%2C%202014 [accessed 2021-12-20]
14. Policy on Systemwide Smoke and Tobacco Free Environment; Executive Order (EO) 1108. The California State University. URL: https://www.csusm.edu/smokefree/pdfs/policy_eo_1108.pdf [accessed 2021-12-19]
15. Mackey TK, Miner A, Cuomo RE. Exploring the e-cigarette e-commerce marketplace: identifying Internet e-cigarette marketing characteristics and regulatory gaps. Drug Alcohol Depend 2015 Nov 01;156:97-103. [doi: 10.1016/j.drugalcdep.2015.08.032]
16. Cuomo RE, Miner A, Mackey TK. Pricing and sales tax collection policies for e-cigarette starter kits and disposable products sold online. Drug Alcohol Rev 2016 Jan;35(1):110-114. [doi: 10.1111/dar.12353] [Medline: 26492956]
17. California School Campus Database. URL: https://www.californiaschoolcampusdatabase.org/ [accessed 2021-12-20]
18. Cuomo RE, Purushothaman VL, Li J, Bardier C, Nali M, Shah N, et al. Characterizing self-reported tobacco, vaping, and marijuana-related tweets geolocated for California college campuses. Front Public Health 2021 Apr 13;9:628812. [doi: 10.3389/fpubh.2021.628812]
19. Cuomo RE, Yang J, Purushothaman VL, Nali M, Li J, Mackey TK. A geospatial analysis of age disparities in resolute localities of tobacco and vaping-specific storefronts in California. Tob Prev Cessat 2021 May;7:32. [doi: 10.18332/tpc/133933]
20. Weber R. Basic Content Analysis (2nd edition). Newbury Park, CA: Sage Publications Inc.; 1990.
21. ATLAS.ti-Scientific Software Development GmbH. Implisense GmbH. URL: https://implisense.com/en/companies/atlas-ti-scientific-software-development-gmbh-berlin-DEWMMDZT3A64 [accessed 2021-12-20]
22. Thomas DR. A general inductive spproach for snalyzing qualitative evaluation data. American Journal of Evaluation 2016 Jun 30;27(2):237-246. [doi: 10.1177/1098214005283748]
23. Ickes MJ, Hahn EJ, McCann M, Kercsmar S. Tobacco-free Take Action!: increasing policy adherence on a college campus. World Med Heal policy 2013;5(1):47-56. [doi: 10.1002/wmh3.2027]
24. Ickes MJ, Rayens MK, Wiggins AT, Hahn EJ. A tobacco-free campus ambassador program and policy compliance. J Am Coll Health 2015;63(2):126-133. [doi: 10.1080/07448481.2014.990972]
25. Harris KJ, Steams JN, Kovach RG, Harrar SW. Enforcing an outdoor smoking ban on a college campus: effects of a multicomponent approach. J Am Coll Health 2009;58(2):121-126. [doi: 10.1080/07448480903221285]

## Abbreviations

**CSU:** California State University
**UC:** University of California



XSL•FO
RenderX

*Edited by A Mavragani; submitted 07.09.21; peer-reviewed by R Dekova, R Krukowski; comments to author 14.10.21; revised version received 07.12.21; accepted 08.12.21; published 24.12.21.*

*Please cite as:*
*Yang JS, Cuomo RE, Purushothaman V, Nali M, Shah N, Bardier C, Obradovich N, Mackey T*
*Campus Smoking Policies and Smoking-Related Twitter Posts Originating From California Public Universities: Retrospective Study*
*JMIR Form Res 2021;5(12):e33331*
*URL: https://formative.jmir.org/2021/12/e33331*
*doi:10.2196/33331*
*PMID:34951597*

©Joshua S Yang, Raphael E Cuomo, Vidya Purushothaman, Matthew Nali, Neal Shah, Cortni Bardier, Nick Obradovich, Tim Mackey. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Transitioning From In-Person to Remote Clinical Research on Depression and Traumatic Brain Injury During the COVID-19 Pandemic: Study Modifications and Preliminary Feasibility From a Randomized Controlled Pilot Study

Lauren B Fisher[1,2], PhD; Sylvie Tuchman[1], BA; Andrew J Curreri[3], MA; Maggie Markgraf[1], BS; Maren B Nyer[1,2], PhD; Paolo Cassano[1,2], MD, PhD; Grant L Iverson[4,5,6], PhD; Maurizio Fava[1,2], MD; Ross D Zafonte[4,5,6,7], DO; Paola Pedrelli[1,2], PhD

[1]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States
[2]Department of Psychiatry, Harvard Medical School, Boston, MA, United States
[3]Department of Psychological & Brain Sciences, Boston University, Boston, MA, United States
[4]Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, MA, United States
[5]Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Boston, MA, United States
[6]Red Sox Foundation and Massachusetts General Hospital Home Base Program, Boston, MA, United States
[7]Department of Physical Medicine and Rehabilitation, Massachusetts General Hospital and Brigham and Women's Hospital, Boston, MA, United States

**Corresponding Author:**
Lauren B Fisher, PhD
Department of Psychiatry
Massachusetts General Hospital
1 Bowdoin Square
6th Floor
Boston, MA, 02114
United States
Phone: 1 617 643 0877
Email: fisher.lauren@mgh.harvard.edu

## Abstract

**Background:** Telehealth has provided many researchers, especially those conducting psychosocial research, with the tools necessary to transition from in-person to remote clinical trials during the COVID-19 pandemic. A growing body of research supports the effectiveness of telemental health for a variety of psychiatric conditions, but few studies have examined telemental health for individuals with comorbid medical diagnoses. Furthermore, little is known about the remote implementation of clinical trials examining telemental health interventions.

**Objective:** This paper outlines the procedural modifications used to facilitate conversion of an in-person randomized controlled trial of cognitive behavioral therapy (CBT) for depression in individuals with traumatic brain injury (TBI; CBT-TBI) to a telemental health study administered remotely.

**Methods:** Given the nature of remote implementation and specific challenges experienced by individuals with TBI, considerations related to treatment delivery, remote consent, data management, neuropsychological assessment, safety monitoring, and delivery of supportive material have been discussed. Feasibility, acceptability, and safety were evaluated by examining attendance and participant responses on self-report measures of treatment satisfaction and suicidal behavior.

**Results:** High rates of treatment attendance, assessment completion, study retention, and satisfaction with the intervention and modality were reported by participants who completed at least one telemental health CBT-TBI session.

**Conclusions:** Study modifications are necessary when conducting a study remotely, and special attention should be paid to comorbidities and population-specific challenges (eg, cognitive impairment). Preliminary data support the feasibility, acceptability, and safety of remotely conducting a randomized controlled trial of CBT-TBI.

**Trial Registration:** ClinicalTrials.gov NCT03307070; https://clinicaltrials.gov/ct2/show/NCT03307070



XSL•FO
RenderX

*(JMIR Form Res 2021;5(12):e28734)* doi:10.2196/28734

**KEYWORDS**

COVID-19; telemental health; clinical trial; traumatic brain injury; depression; cognitive behavioral therapy

## Introduction

The COVID-19 pandemic has had significant impacts on the conduct of clinical research. As a result of stay-at-home orders, most in-person clinical trials for nonlifesaving interventions were suddenly halted, leaving many active research participants and teams in limbo. Many researchers were forced to choose between pausing their research projects, with a threat to scientific productivity, and modifying procedures to implement remote approaches [1].

Psychosocial research (eg, investigation of psychological treatments) is well positioned to be conducted remotely. Telemental health, the application of telecommunications to provide mental health services from a distance, has grown exponentially during the COVID-19 pandemic [2] and includes the use of a wide range of technologies to deliver synchronous (eg, live videoconferencing or telephone calls) and asynchronous interventions (eg, web-based interventions completed without a clinician present) [3]. This paper primarily focuses on the use of synchronous exchanges with telephone and videoconferencing to facilitate the remote implementation of assessment and psychotherapy in the context of a clinical trial for depression that was conducted in-person prior to the pandemic. The transition to remote study implementation was supported by a growing body of research demonstrating the effectiveness of telemental health services across many populations (eg, adults, children, and older adults) and for a range of psychiatric conditions [4], including depression [5], anxiety [6], and posttraumatic stress disorder [7]. In fact, real-time telemental health (ie, videoconferencing or telephone) is as effective as face-to-face treatment in reducing depressive symptoms [5,8]. Furthermore, treatment satisfaction and therapeutic alliance are similar among patients engaged in telemental health (videoconferencing and telephone-based interventions) and in-person treatment [9]. Effective implementation of protocols for remotely assessing and managing suicide risk further supports the feasibility of conducting clinical trials that examine telemental health interventions for individuals with depression [10].

Despite the significant promise of telemental health for many individuals with depression, individuals with various comorbid medical diagnoses may experience distinct challenges that serve as barriers to effectively utilizing telemental health interventions and participating in clinical research implemented remotely. In our work with individuals who have sustained traumatic brain injury (TBI), the impacts of TBI sequelae, including cognitive difficulties (eg, impaired focus and attention, and executive dysfunction) and sensitivity to light/screens, present unique challenges to participation in telemental health. Nevertheless, preliminary evidence suggests that individuals with major depressive disorder (MDD) and complicated mild to severe TBI experience similar reductions in depressive symptoms after 16 weeks of in-person and telephone-delivered cognitive behavioral therapy (CBT), and report similarly high rates of treatment satisfaction and strong therapeutic alliance [11]. Although we have been unable to identify studies examining the use of videoconferencing for the delivery of individual psychotherapy for adults with depression and TBI, there is support for the feasibility of using videoconferencing for group CBT to improve emotion regulation after TBI [12] and problem solving–based telemental health for improving behavior and family functioning in children with TBI [13]. Furthermore, support for the feasibility of video-based telerehabilitation in adults with TBI [14] suggests that telemental health interventions adapted for this population may also be feasible.

Given the mounting evidence for telemental health research as a feasible alternative to face-to-face participation in clinical trials and the public health restrictions during the COVID-19 pandemic, in March 2020, we transitioned our National Institutes of Health (NIH)-funded, in-person randomized controlled trial (RCT) of CBT for depression in individuals with TBI (CBT-TBI) to a remotely delivered telemental health study. In this paper, we discuss the study modifications that were implemented to maximize the feasibility, acceptability, and safety of remote study participation for individuals with depression and TBI. Preliminary evidence of feasibility, acceptability, and safety is examined among individuals who began the study in-person and transitioned to remote procedures, as well as participants who completed all procedures remotely.

## Methods

### Participants

Study participants were enrolled in an ongoing randomized waitlist-controlled trial (target N=40) piloting a 12-week individual CBT for depression that was adapted for individuals with TBI. The aims of the parent trial were to evaluate the feasibility and acceptability of the intervention (primary), as well as the potential efficacy in reducing depressive severity (secondary). As of August 27, 2021, the ongoing RCT enrolled (consented) a total of 33 participants, of which 18 participants were enrolled in-person and 15 participants were enrolled remotely. The clinical trial began in-person recruitment in April 2019 and transitioned to remote procedures on March 16, 2020, which remains ongoing at the time of writing. Of the 33 enrolled participants, 8 completed all study visits in-person, 6 completed a combination of in-person and telemental health sessions, 2 were enrolled in-person but completed all CBT-TBI sessions remotely, 7 completed all study visits remotely, 3 were deemed ineligible at the screening session (1 in-person and 2 remote), 2 discontinued (during remote CBT-TBI), 1 was lost to follow-up, and 3 determined they did not have time to participate (immediately after remote enrollment). One participant remains active at the time of writing. The feasibility and safety analyses presented below include participants who completed one or more CBT-TBI sessions remotely. Acceptability data include participants who completed one or more CBT-TBI sessions



XSL•FO
RenderX

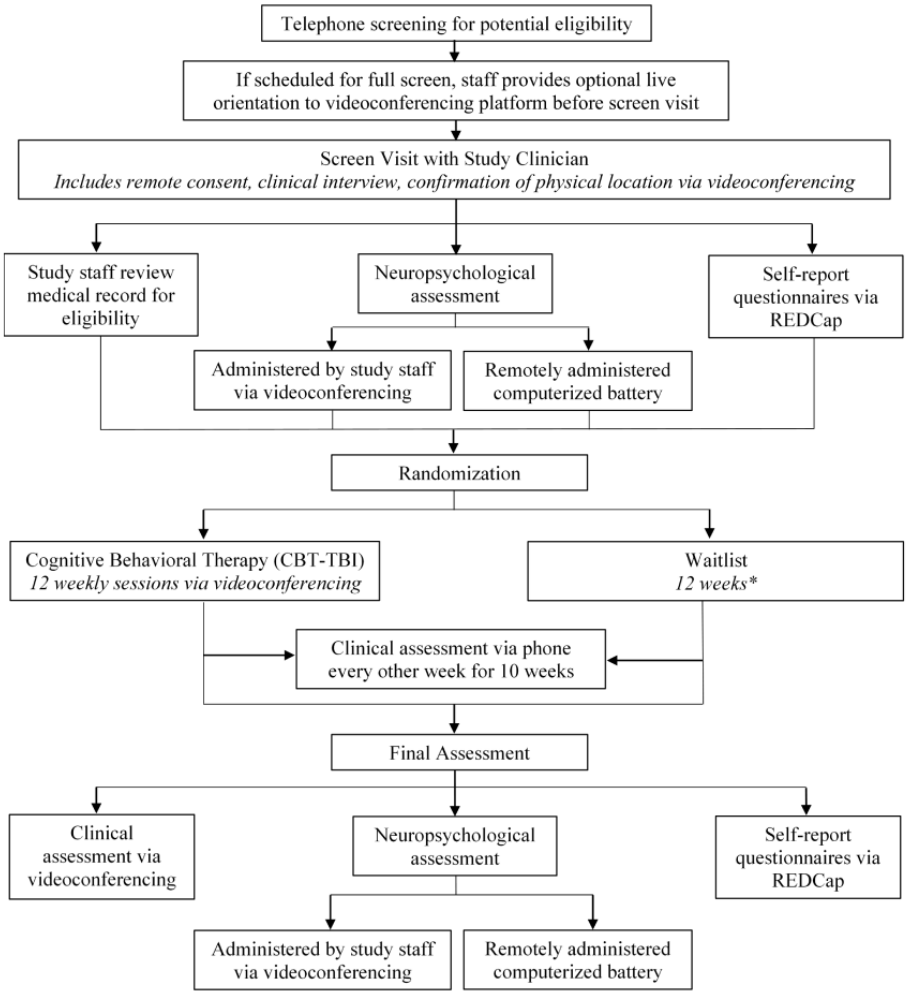
remotely and completed the study (defined as attending all 12 intervention sessions; n=12), as well as 1 participant who terminated CBT-TBI early but completed end-of-study assessments. Participants in this subsample of the ongoing RCT (n=18) were between the ages of 21 and 69 years (mean 39.2 years, SD 16.3 years), and the majority were white (n=14, 78%), non-Hispanic or Latino (n=15, 83%), and highly educated (n=10, 56% with at least 4 years of college). Just over half of the sample (n=10, 56%) were women, while less than half the sample (n=8, 44%) were married or in a relationship.

## Procedures

All study procedures, including pandemic-related modifications, were approved by the Massachusetts General Hospital Institutional Review Board. Adults with clinically significant depressive symptoms and a history of moderate-to-severe TBI were included in the study (see Multimedia Appendix 1 for the full study criteria [15-18]).

The initial screening visit included informed consent, diagnostic and symptom evaluation with a study clinician, and a neuropsychological battery that was completed in-person prior to March 2020 and remotely since April 2020. Participants also completed a series of baseline self-report measures of mood; suicidality; and cognitive, social, and emotional functioning, using a secure web-based platform (REDCap). Eligible participants were then randomized to 12 weeks of a newly developed manualized cognitive behavioral treatment for depression adapted for individuals with TBI (CBT-TBI) or a 12-week waitlist. The intervention included psychoeducation, behavioral activation, goal setting, cognitive restructuring, and relapse prevention. CBT-TBI was adapted for individuals with TBI by incorporating the following strategies: repetition, patient workbook with session summaries and forced choice worksheets, modified thought records, therapeutic use of neuropsychological testing results, individually tailored text messages/between-session reminders, and daily use of an activity monitoring device (Fitbit Charge 3). Individuals in both conditions completed bimonthly phone assessments of depressive symptoms with an independent evaluator. Weekly 50 to 60-minute individual CBT-TBI sessions were delivered by a master's or doctoral-level clinician in-person until March 16, 2020, and via Zoom videoconferencing (or telephone, when needed) thereafter. At the end of 12 weeks, all participants completed a postassessment, which included clinician-rated and self-rated measures, as well as repeat neuropsychological testing. Individuals randomized to the waitlist condition could receive CBT-TBI upon completion of 12 weeks of assessment. All study procedures are outlined in Figure 1, and a detailed breakdown of CBT-TBI visits is shown in Figure 2.

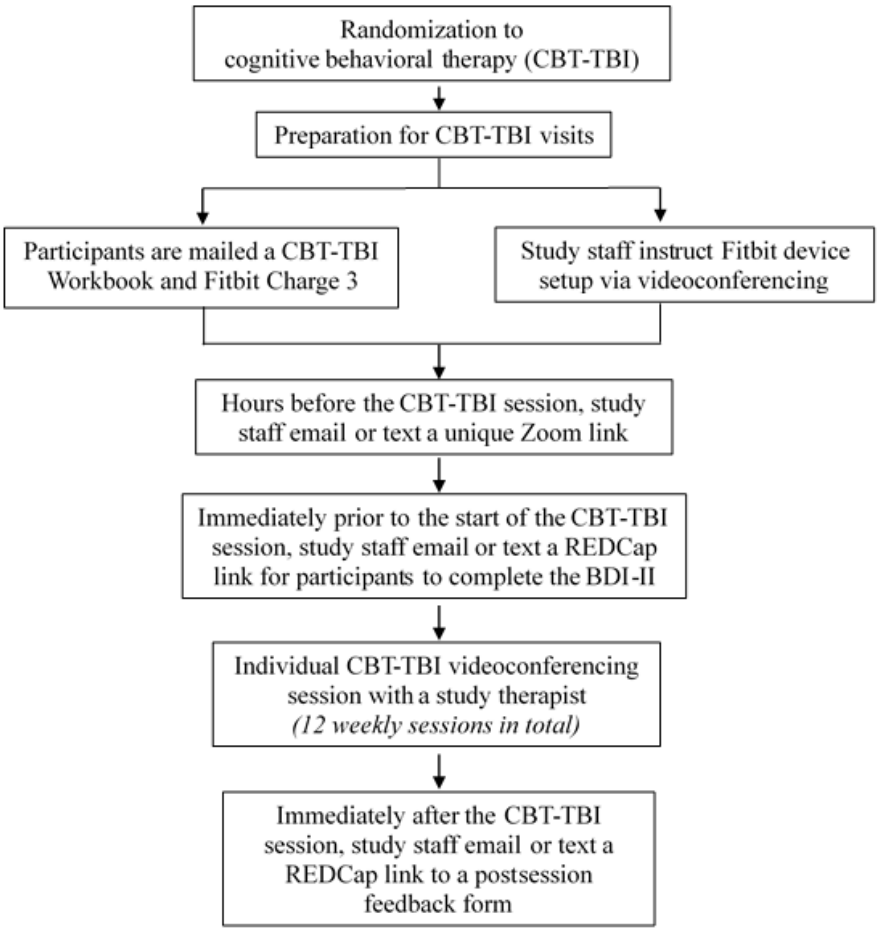
**Figure 1.** Flowchart of study procedures for eligible participants. *All participants randomized to waitlist can complete the intervention following the final assessment. CBT-TBI: cognitive behavioral therapy for depression in individuals with traumatic brain injury.






XSL•FO
RenderX

**Figure 2.** Flowchart of study procedures for treatment phase. BDI-II: Beck Depression Inventory-II; CBT-TBI: cognitive behavioral therapy for depression in individuals with traumatic brain injury.




## Measures

### *Acceptability*

The 12-item self-rated Satisfaction with Therapy and Therapist Scale-Revised (STTS-R) [19] was used to assess satisfaction in 2 domains of treatment. Current analyses included the satisfaction with therapy subscale scores, which range from 6 to 30, with higher scores indicating greater satisfaction. In March 2020, 5 questions were composed by the study team to gather feedback on remote CBT-TBI visits, including satisfaction with remote CBT-TBI sessions on a 5-point Likert scale (1, very satisfied to 5, very dissatisfied). Participants were also asked to share what they liked and did not like about virtual treatment. Participants who completed some CBT-TBI sessions in-person and some remotely were asked to indicate the degree of their preference for one modality over the other on a 5-point Likert scale (1, strongly preferred telemental health sessions to 5, strongly preferred in-person sessions). Finally, participants were asked to select the modality they would choose if given the option for treatment after the pandemic (eg, in-person, over the telephone, videoconferencing, and combination of in-person and virtual).

### *Safety*

Suicidal ideation was monitored weekly during CBT-TBI with the suicide item from the Beck Depression Inventory-II (BDI-II) [20], a 21-item self-report scale designed to measure the presence and severity of depressive symptoms. The BDI-II suicide item is associated with the risk of repeat suicide attempts and death by suicide and is recommended as a screener for suicide risk in routine clinical care [21]. Adverse events were also assessed during bimonthly phone assessments.

## Study Modifications With Transition to Remote Implementation of Research

Several protocol modifications (Table 1) were instituted after all clinical trials for nonlifesaving interventions were halted in our institution due to the pandemic. Modifications aimed to facilitate feasibility and adherence to the original procedures as much as possible.



XSL•FO
RenderX

**Table 1.** Study modifications with transition to remote implementation.

| Protocol element | In-person implementation | Remote implementation |
|---|---|---|
| Treatment modality | Individual face-to-face sessions in the office | Individual sessions via secure videoconferencing |
| Consent | Clinicians and participants review paper consent and sign in the office | Clinicians and participants utilize teleconsent with the REDCap eConsent template during videoconference |
| Data management | • Participants complete questionnaires directly on REDCap using an in-office computer (preferred method)<br>  • REDCap links are emailed to participants who are unable to complete questionnaires during office visit<br>  • Paper copies are completed in office or at home for participants unable to complete electronically<br>• Clinicians complete pencil and paper assessments (requires data entry) | • Via text or email, participants are sent a REDCap link to complete questionnaires independently (preferred method)<br>  • Paper copies are mailed with self-addressed return envelope for participants unable to complete electronically<br>• Clinicians enter clinical data directly into REDCap |
| Neuropsychological assessment | • Administered by study staff in the office (traditional measures):<br>  • TOPF[a] [22]<br>  • WAIS-IV[b] Coding, Digit Span and Similarities [23]<br>  • D-KEFS[c] Color Word and Trails [24]<br>  • CVLT-II[d] [25]<br>• Administered on an iPad in the office:<br>  • NIH[e] Toolbox Cognition Battery [26] | • Administered by study staff via videoconferencing (traditional measures)<br>  • TOPF [22]<br>  • WAIS-IV Digit Span and Similarities<br>  • WMS-IV[f] Logical Memory I and II [27]<br>• Computerized battery completed by participants at home<br>  • CNS Vital Signs [28] |
| Suicide risk monitoring | Clinicians review the paper copy responses to the BDI-II[g] suicide item at the start of the CBT-TBI[h] visit with participants in the room | The study coordinator reviews REDCap responses to the BDI-II suicide item at the start of the CBT-TBI visit and alerts clinicians to scores of 2 or higher |
| Preparation for CBT-TBI visits | Routine scheduling; the study coordinator answers questions from participants | • A "Welcome Letter" is sent to establish expectations:<br>  • Ensure security (eg, close other applications while Zoom is open)<br>  • Ensure privacy (eg, conduct sessions in a private room with the door closed, use headphones and/or a noise blocker)<br>  • Provides tips for limiting distractions (eg, silence cell phone, avoid eating, ensure device is fully charged, and let others in the home/space know you are unavailable)<br>  • Consider the feasibility of your device (eg, a computer allows for typing notes in electronic handouts and hardwired ethernet connections can be more reliable than Wi-Fi) |
| Delivery and setup of wearable technology | The study coordinator sets up Fitbit with participants on the day of the first in-office CBT-TBI session | The study coordinator mails the device to participants and guides them through device setup via videoconferencing |
| CBT-TBI delivery modification | Provide handouts in the session that are added to the CBT-TBI Workbook every week | • Mail the CBT-TBI Workbook with handouts and worksheets prior to the start of treatment<br>• Minimize reliance on screens (eg, turn away from the computer and turn off video)<br>• Tailor delivery to individual needs/preferences and be flexible (eg, utilize "screen share" and provide electronic handouts) |

[a]TOPF: Test of Premorbid Functioning.

[b]WAIS-IV: Wechsler Adult Intelligence Scale–Fourth Edition.

[c]D-KEFS: Delis-Kaplan Executive Function System.

[d]CVLT-II: California Verbal Learning Test–Second Edition.



XSL•FO
RenderX

[e]NIH: National Institutes of Health.

[f]WMS-IV: Wechsler Memory Scale–Fourth Edition.

[g]BDI-II: Beck Depression Inventory-II.

[h]CBT-TBI: cognitive behavioral therapy for depression in individuals with traumatic brain injury.

### Videoconferencing Platform

All procedures (except for neuropsychological testing, discussed below) that were previously conducted face-to-face were completed remotely using videoconferencing. Secure Zoom videoconferencing (Zoom Enterprise) was adopted after working with our institution's Research Information Security Office to optimize privacy and security settings, including enabling the waiting room, locking meetings once sessions begin, and generating meeting IDs with a password. Individuals who were unfamiliar with the videoconferencing platform received step-by-step instructions for Zoom account set-up, and could participate in a "trial run" and orientation to the platform with the study coordinator.

### Remote Consent

Since the study transitioned to remote implementation, the informed consent process was embedded into a live telehealth session, also referred to as teleconsent [29]. An institution-specific REDCap electronic informed consent template was utilized. The study clinician met with the participant over Zoom to review the consent form and instruct the participant to digitally sign consent. A signed copy was then securely emailed directly to the participant from REDCap. Participants were given the option to receive a mailed paper copy and/or a brief summary of key study information to make the process less overwhelming. Overall, teleconsent provides a feasible alternative to in-person paper consent and facilitates research continuity when face-to-face interactions are not possible [30].

### Data Management

Prior to the pandemic, study participants were given the option of completing self-report questionnaires directly in REDCap using a computer or tablet, either in our office or at home. Individuals participating remotely were provided an electronic link to complete questionnaires at home directly on REDCap. Given that it can be cognitively taxing to sit at a screen for an extended duration of time, participants were encouraged to complete a few questionnaires at a time. Participants were offered the option of being mailed paper questionnaires with a self-addressed envelope that was returned to the research team.

### Neuropsychological Testing

The original in-person battery included a series of traditional paper and pencil neuropsychological measures and the iPad-administered NIH Toolbox Cognition battery [26]. Following a review of the available teleneuropsychology literature [31] and guidelines for remote assessments [32], it was determined that certain subtests from the original battery could be administered via videoconferencing, likely without significant impact on reliability and validity, although some tests could not (Table 1). CNS Vital Signs [28], a brief computerized neurocognitive test battery, replaced the NIH Toolbox Cognition battery and was administered remotely in accordance with guidelines to maximize validity. Participants were instructed to watch a preparatory video that emphasizes the importance of creating an optimal standardized testing environment (eg, limit distractions and interruptions, and set aside sufficient time to complete), which in turn maximizes the reliability of test results. CNS Vital Signs reports include a validity indicator for each subtest, allowing the clinician to follow-up with the participant about possibly invalid results.

### Suicide Risk Assessment

On the day of CBT-TBI sessions, the study coordinator emailed or texted the REDCap link for the BDI-II [20] to the participants, instructed them to complete the measure prior to meeting with the study therapist, and reviewed their responses to the BDI-II suicide item in real-time. If the participant did not complete the measure, the study therapist was notified to remind the participant to do so and to review the suicide item response before starting the session. In the event that the participant had a score of 2 or higher, the therapist was immediately alerted to conduct a detailed suicide risk assessment in the session and to determine the need for a higher level of care, which could involve voluntary or involuntary hospitalization and/or contacting the individual's previously identified emergency contact. The participant's physical location was identified after signing consent and was confirmed before starting every CBT-TBI session. All efforts were made for the study therapist to remain connected to the participant (on Zoom or telephone) until emergency personnel arrived at their location.

### Preparing for CBT-TBI Telehealth Visits

Given that individuals with TBI can be sensitive to changes in routine due to deficits in mental flexibility and problem solving [33], it is encouraged that telemental health visits mirror in-person CBT-TBI visits as much as possible. A predictable environment that parallels the in-person setting (ie, consistent office space/background) may be beneficial [34]. In order to compensate for a patient's reduced ability to read the therapist's nonverbal cues over video, clinicians configured their camera to ensure that the patient could see as much of their body language as was feasible [34]. Clear and consistent expectations about virtual visits were directly communicated in a "Welcome Letter" that emphasized how best to ensure security and privacy, and provided tips for limiting distractions during the session (Table 1). Participants who had more than one internet and video-enabled device were encouraged to consider which device would best suit their needs based on factors such as strength of internet connection and device portability. Additionally, the active hands-on nature of the study intervention, which uses worksheets and encourages notetaking, warranted an appropriate workspace, such as sitting at a desk or table.

### Delivery and Setup of Wearable Technology

Participants were mailed a Fitbit Charge 3 activity tracker prior to starting the intervention. The study coordinator scheduled individual videoconferencing meetings with participants prior



XSL•FO
RenderX

to the first CBT-TBI session where they provided instruction on setup and use, and provided a handout reiterating this information.

### *Minimize Reliance on Screens*

For some individuals with TBI, screen time can exacerbate symptoms, including headaches due to photosensitivity [35]. To reduce the degree to which participants engaged with material in electronic format, all participants were mailed a physical copy of the CBT-TBI Workbook, which contained copies of weekly agendas, handouts, and worksheets. Participants who reported considerable difficulty with photosensitivity were encouraged to turn away from the computer or turn off video.

### *Importance of Tailoring Delivery and Being Flexible*

Given the importance of clear direct communication for individuals experiencing cognitive difficulties, the therapist spoke with the participants in the first session to understand their comfort with technology and preferences for ways of engaging in the collaborative treatment (eg, physical vs electronic worksheets, therapist vs participant typing responses into worksheets, and use of the videoconferencing screen share feature to provide visual cues and allow the therapist to model skill use). Consistent with procedures utilized during in-person delivery of the intervention, participants continued to receive between-session reminders via text, email, or phone call to carry out collaboratively identified activities or goals (eg, behavioral activation).

### *Troubleshooting Challenges With Technology*

It is inevitable that technological challenges will arise both prior to and during virtual sessions. It is important that clinicians do not get visibly frustrated in the face of technological difficulties, as the patient may interpret this as the clinician being upset with them [34]. Additionally, shared insecurities over technology may, in fact, aid therapeutic alliance [36]. However, it is worth recognizing that technological issues can be disruptive, and clinicians may want to identify a cutoff point at which they switch from videoconferencing to a telephone session. In the first CBT-TBI session, the clinician and participant develop an individualized plan for navigating potential technological difficulties, such as losing a connection mid-session.

## Data Analyses

To assess the preliminary feasibility and acceptability of remote study implementation, descriptive statistics were used to report the number of CBT-TBI intervention sessions attended, number of assessment sessions attended, number of participants who completed the study, and rate of satisfaction with treatment (STTS-R and supplemental questions). Study retention (number of study completers/number randomized) was also calculated. Finally, the number of elevated responses to the BDI-II suicide item (≥2) was described to demonstrate the maintenance of safety protocols.

# Results

## Feasibility

At the time of the transition to remote procedures (March 2020), there were 9 active study participants, including 3 participants on the waitlist (ie, had not started CBT-TBI) and 6 participants who were mid-treatment. The 6 participants who started CBT-TBI in-person prior to March 2020 were at different points in treatment at the time of the transition (weeks 3, 7, 9, 10 11, and 12), and all went on to complete the remainder of their 12 weeks of CBT-TBI remotely using telemental health sessions. Two out of three waitlist participants who enrolled in the study before March 2020 with the expectation of attending CBT-TBI sessions in-person completed the entire 12 weeks of CBT-TBI via videoconferencing. One participant discontinued participation after 4 CBT-TBI sessions due to a demanding work schedule but completed postassessments. Finally, of all randomized participants who enrolled in the study remotely (n=9), 7 (78%) completed the study, 1 (11%) was withdrawn due to worsening depression, and 1 (11%) remains active in CBT-TBI. The study retention rate prior to March 2020 was 100% (8 CBT-TBI completers), and completion from March through January 2021 was about 93% (13 out of 14 possible randomized CBT-TBI completers). Approximately 91% of clinician-rated assessments (102 out of 112 possible assessments) were completed throughout the period in which the study had been conducted remotely.

## Acceptability

Satisfaction with therapy (STTS-R therapy subscale) was high (mean 27.1, SD 2.8) among participants who completed at least one session of CBT-TBI remotely (n=16). Overall satisfaction with telemental health sessions (videoconferencing, telephone, or a combination of both) was high (n=14); 9 participants (64%) reported being “very satisfied,” 4 participants (29%) were “satisfied,” and 1 participant (7%) reported being “neither satisfied nor dissatisfied.” If given a choice of modality in the future (n=14), 3 participants (21%) indicated that they would choose in-person treatment, 4 participants (29%) indicated that they would choose telemental health treatment (videoconferencing or telephone), and 7 participants (50%) indicated that they would choose a combination of in-person and telemental health treatment. Among the participants who completed a combination of in-person and telemental health treatment and provided feedback (n=4), there was no clear pattern in preferred modality, as they reported strongly preferring telemental health treatment (n=1, 25%), somewhat preferring telemental health treatment (n=1, 25%), strongly preferring in-person treatment (n=1, 25%), and no indication of preference (n=1, 25%).

Qualitative feedback highlighted that all 14 study completers noted at least one benefit of telemental health sessions, including ease of conducting sessions from home and not having to travel for appointments. Conversely, technological challenges, reduced focus, limited privacy, and difficulty feeling connected with the therapist were noted as factors that participants disliked about telemental health sessions. Five participants reported that there was nothing they disliked about telemental health treatment.



XSL•FO
RenderX

### Safety

Since March 2020, 1 participant had a score of 3 on the BDI-II [20] suicide item in 2 consecutive weeks. Per protocol, the study coordinator alerted the CBT-TBI therapist and the principal investigator immediately after having identified the safety concern, and the therapist started the CBT-TBI session with a thorough assessment of suicide risk. After several weeks of worsening depression and increasing suicidal ideation, the participant was eventually referred for a higher level of care (partial hospitalization program), was withdrawn from the study, and was ultimately hospitalized voluntarily for worsening of symptoms. No serious adverse events were reported throughout the duration of remote procedures.

## Discussion

### Principal Findings

Using several procedural modifications described in this paper, an in-person RCT of CBT for depression after TBI was converted to remote implementation and demonstrated preliminary evidence of feasibility, acceptability, and safety. Specific modifications to study implementation and the treatment protocol are outlined in Table 1. Given the range of symptoms and deficits that can arise after TBI (eg, photosensitivity and impaired attentional capacity), all modifications were made with consideration of their potential impact on participants and to enhance feasibility.

Preliminary data supported the feasibility, acceptability, and safety of conducting an RCT for depression among individuals with TBI exclusively utilizing remote procedures. Specifically, preliminary results demonstrated a high rate of completion for clinician-rated assessments (102/112, 91%) and high study retention (13/14, 93%). Procedures that were designed to monitor safety were effective in identifying individuals at high risk for suicide, triggering clinician suicide risk assessments via videoconferencing. Feedback from participants suggested a high degree of satisfaction with the CBT-TBI treatment and telemental health modality, providing initial evidence of the acceptability of the remotely delivered study intervention. The findings are consistent with the results of previous studies that have examined telephone-delivered cognitive behavioral interventions among individuals with TBI [11,37].

Feedback from our small sample highlighted a range of preferences when participants were asked to consider their ideal treatment modality (in-person treatment, telemental health treatment, or a hybrid model), which has significant implications for study participation and potentially for treatment outcomes. Research has demonstrated better treatment outcomes among individuals whose preferences about psychological treatment (eg, appointment time, venue, and treatment type) are accommodated compared with individuals whose preferences are not met [38]. Previous research among depressed individuals with TBI utilized choice-stratified randomization, in which participants could assert a preference for CBT that was delivered in-person or over the telephone prior to randomization, in order to enhance ecological validity [11]. Qualitative feedback from our study suggested that participants may appreciate a mix of in-person and telemental health visits, which is consistent with evidence for the high feasibility and acceptability of “blended” models of delivery (combination of face-to-face and web-based sessions) of CBT for depression [39]. Although the efficacy of our study intervention is unknown at this time, tailoring the intervention modality according to preferences may lead to greater attendance at treatment sessions and engagement in treatment.

### Limitations

It is important to acknowledge several limitations. Although several steps were taken to optimize the testing environment, neuropsychological assessment is ideally suited to in-person administration. Challenges with technology and suboptimal conditions in the participant’s environment have the potential to impact engagement and data collection. Behavioral observations can be restricted by videoconferencing, and rapport can sometimes be limited without in-person interactions, which may impact participant responses or commitment to participation, especially prior to randomization. It is also important to note that our sample was heavily comprised of individuals who received specialized acute inpatient rehabilitation (n=11, 61%) and specialized outpatient treatment (n=13, 72%) for their TBIs in a single academic medical center in the Northeast. Individuals who receive inpatient rehabilitation represent 7% of all persons hospitalized with moderate-to-severe TBI, are less likely to be a member of a racial/ethnic minority group, and are more likely to have health insurance compared with individuals who are hospitalized and do not receive inpatient rehabilitation after moderate-to-severe TBI [40]. Thus, our sample may not be representative of all individuals with moderate-to-severe TBI in the United States.

### Conclusion

Remote study participation has been a feasible alternative when in-person research was halted during the COVID-19 pandemic. Strategic procedural modifications outlined in this paper have been instrumental to the continued feasibility of recruitment and retainment of individuals with depression and TBI in the context of our ongoing RCT. Furthermore, telemental health offers significant advantages in eliminating common barriers to study participation, including transportation, time needed to travel to appointments, distance to the hospital, limited mobility, and inclement weather. Conversely, some individuals may struggle to secure private space for their sessions, and a visit to a traditional office space may be preferred. Further, many individuals do not have internet access and a camera-enabled device for videoconferencing. Some individuals may find it easier and less intimidating to be vulnerable about the challenges they face over a computer screen rather than in-person [41], while others may have difficulty connecting with a therapist through a screen. For some individuals, the flexibility of utilizing both types of modalities within the course of treatment may be an ideal balance; thus, future research designs should consider the role of patient preference. For individuals with TBI who frequently struggle with physical, cognitive, and emotional impairments, flexibly tailored treatments that utilize telemental health and in-person modalities are likely to be important in both research and clinical settings. Future research should directly compare the feasibility and efficacy of CBT delivered



XSL•FO
RenderX

via telemental health, in-person, and hybrid models for individuals with TBI, as well as the validity and reliability of remote neuropsychological assessments and strategies to facilitate remote engagement. The COVID-19 pandemic abruptly presented researchers with unique challenges that have required flexibility and innovation. The advantages presented by the ability to conduct clinical research using remote methods are likely to persist long after the pandemic ends.

## Acknowledgments

The research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number K23HD087464. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of Interest

PC co-founded and consults for Niraxx Light Therapeutics Inc. He has no conflicts of interest related to this publication. GLI serves as a scientific advisor for NanoDx, Sway Operations, LLC, and Highmark, Inc. He has a clinical and consulting practice in forensic neuropsychology, including expert testimony, involving individuals who have sustained mild TBIs (including athletes). He has received research funding from several test publishing companies, including ImPACT Applications Inc, CNS Vital Signs, and Psychological Assessment Resources (PAR, Inc). He has received research funding as a principal investigator from the National Football League, and salary support as a collaborator from the Harvard Integrated Program to Protect and Improve the Health of National Football League Players Association Members. He acknowledges unrestricted philanthropic support from ImPACT Applications, Inc, the Mooney-Reed Charitable Foundation, the National Rugby League, and the Spaulding Research Institute. These entities were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. RDZ received royalties from Springer/Demos publishing for serving as co-editor of the text Brain Injury Medicine. RDZ serves on the Scientific Advisory Board of Myomo and Onecate.ai. He has no conflicts of interest related to this publication. The other authors have no conflicts relevant to this publication.

Multimedia Appendix 1
Inclusion/exclusion criteria for the randomized controlled trial.
[DOCX File , 30 KB - formative_v5i12e28734_app1.docx ]

Multimedia Appendix 2
CONSORT checklist.
[DOC File , 219 KB - formative_v5i12e28734_app2.doc ]

## References

1. Upadhaya S, Yu JX, Oliva C, Hooton M, Hodge J, Hubbard-Lucey VM. Impact of COVID-19 on oncology clinical trials. Nat Rev Drug Discov 2020 Jun;19(6):376-377. [doi: 10.1038/d41573-020-00093-1] [Medline: 32424343]
2. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of Telehealth During the COVID-19 Pandemic: Scoping Review. J Med Internet Res 2020 Dec 01;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]
3. Barnett JE, Kolmes K. The practice of tele-mental health: Ethical, legal, and clinical issues for practitioners. Practice Innovations 2016;1(1):53-66. [doi: 10.1037/pri0000014]
4. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]
5. Osenbach JE, O'Brien KM, Mishkind M, Smolenski DJ. Synchronous telehealth technologies in psychotherapy for depression: a meta-analysis. Depress Anxiety 2013 Nov 06;30(11):1058-1067. [doi: 10.1002/da.22165] [Medline: 23922191]
6. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009 Jun;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
7. Sloan DM, Gallagher MW, Feinstein BA, Lee DJ, Pruneau GM. Efficacy of telehealth treatments for posttraumatic stress-related symptoms: a meta-analysis. Cogn Behav Ther 2011 Jun;40(2):111-125. [doi: 10.1080/16506073.2010.550058] [Medline: 21547778]
8. Berryhill MB, Culmer N, Williams N, Halli-Tierney A, Betancourt A, Roberts H, et al. Videoconferencing Psychotherapy and Depression: A Systematic Review. Telemed J E Health 2019 Jun;25(6):435-446. [doi: 10.1089/tmj.2018.0058] [Medline: 30048211]
9. Jenkins-Guarnieri MA, Pruitt LD, Luxton DD, Johnson K. Patient Perceptions of Telemental Health: Systematic Review of Direct Comparisons to In-Person Psychotherapeutic Treatments. Telemed J E Health 2015 Aug;21(8):652-660. [doi: 10.1089/tmj.2014.0165] [Medline: 25885491]



XSL•FO
RenderX

10. Luxton DD, Pruitt LD, O'Brien K, Stanfill K, Jenkins-Guarnieri MA, Johnson K, et al. Design and methodology of a randomized clinical trial of home-based telemental health treatment for U.S. military personnel and veterans with depression. Contemp Clin Trials 2014 May;38(1):134-144. [doi: 10.1016/j.cct.2014.04.002] [Medline: 24747488]
11. Fann JR, Bombardier CH, Vannoy S, Dyer J, Ludman E, Dikmen S, et al. Telephone and in-person cognitive behavioral therapy for major depression after traumatic brain injury: a randomized controlled trial. J Neurotrauma 2015 Jan 01;32(1):45-57 [FREE Full text] [doi: 10.1089/neu.2014.3423] [Medline: 25072405]
12. Tsaousides T, D'Antonio E, Varbanova V, Spielman L. Delivering group treatment via videoconference to individuals with traumatic brain injury: a feasibility study. Neuropsychol Rehabil 2014;24(5):784-803. [doi: 10.1080/09602011.2014.907186] [Medline: 24810148]
13. Wade SL, Raj SP, Moscato EL, Narad ME. Clinician perspectives delivering telehealth interventions to children/families impacted by pediatric traumatic brain injury. Rehabil Psychol 2019 Aug;64(3):298-306. [doi: 10.1037/rep0000268] [Medline: 30816734]
14. Ownsworth T, Arnautovska U, Beadle E, Shum D, Moyle W. Efficacy of Telerehabilitation for Adults With Traumatic Brain Injury: A Systematic Review. J Head Trauma Rehabil 2018;33(4):E33-E46. [doi: 10.1097/HTR.0000000000000350] [Medline: 29084100]
15. Sheehan D, Lecrubier Y, Sheehan K, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
16. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger J, Burns C. The inventory for depressive symptomatology (IDS): Preliminary findings. Psychiatry Research 1986 May;18(1):65-87. [doi: 10.1016/0165-1781(86)90060-0]
17. Jeste DV, Palmer BW, Appelbaum PS, Golshan S, Glorioso D, Dunn LB, et al. A new brief instrument for assessing decisional capacity for clinical research. Arch Gen Psychiatry 2007 Aug 01;64(8):966-974. [doi: 10.1001/archpsyc.64.8.966] [Medline: 17679641]
18. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. The Lancet 1974 Jul;304(7872):81-84. [doi: 10.1016/s0140-6736(74)91639-0]
19. Oei TPS, Green AL. The Satisfaction With Therapy and Therapist Scale--Revised (STTS-R) for group psychotherapy: Psychometric properties and confirmatory factor analysis. Professional Psychology: Research and Practice 2008 Aug;39(4):435-442. [doi: 10.1037/0735-7028.39.4.435]
20. Beck A, Steer R, Brown G. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.
21. Green KL, Brown GK, Jager-Hyman S, Cha J, Steer RA, Beck AT. The Predictive Validity of the Beck Depression Inventory Suicide Item. J Clin Psychiatry 2015 Dec;76(12):1683-1686. [doi: 10.4088/JCP.14m09391] [Medline: 26717528]
22. Advanced clinical solutions for WAIS-IV and WMS-IV: Administration and scoring manual. San Antonio, TX: Psychological Corporation; 2009.
23. Wechsler D. WAIS-IV: Administration and scoring manual. San Antonio, TX: Psychological Corporation; 2008.
24. Delis D, Kaplan E, Kramer J. Delis-Kaplan Executive Function System (D-KEFS): Examiner's Manual. San Antonio, TX: NCS Pearson, Inc; 2001.
25. Delis D, Kramer J, Kaplan E, Ober B. California Verbal Learning Test - Second Edition (CVLT-II). San Antonio, TX: Psychological Corporation; 2000.
26. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, et al. Cognition assessment using the NIH Toolbox. Neurology 2013 Mar 12;80(11 Suppl 3):S54-S64 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872ded] [Medline: 23479546]
27. Wechsler D. Wechsler Memory Scale - Fourth Edition (WMS-IV): Administration and scoring manual. San Antonio, TX: Pearson; 2009.
28. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. Arch Clin Neuropsychol 2006 Oct;21(7):623-643. [doi: 10.1016/j.acn.2006.05.007] [Medline: 17014981]
29. Welch BM, Marshall E, Qanungo S, Aziz A, Laken M, Lenert L, et al. Teleconsent: A Novel Approach to Obtain Informed Consent for Research. Contemp Clin Trials Commun 2016 Aug 15;3:74-79 [FREE Full text] [doi: 10.1016/j.conctc.2016.03.002] [Medline: 27822565]
30. Skelton E, Drey N, Rutherford M, Ayers S, Malamateniou C. Electronic consenting for conducting research remotely: A review of current practice and key recommendations for using e-consenting. Int J Med Inform 2020 Nov;143:104271 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104271] [Medline: 32979650]
31. Marra DE, Hamlet KM, Bauer RM, Bowers D. Validity of teleneuropsychology for older adults in response to COVID-19: A systematic and critical review. Clin Neuropsychol 2020 Jun 10;34(7-8):1411-1452. [doi: 10.1080/13854046.2020.1769192] [Medline: 32519594]
32. Bilder RM, Postal KS, Barisa M, Aase DM, Cullum CM, Gillaspy SR, et al. InterOrganizational practice committee recommendations/guidance for teleneuropsychology (TeleNP) in response to the COVID-19 pandemic. Clin Neuropsychol 2020 Jul 16;34(7-8):1314-1334 [FREE Full text] [doi: 10.1080/13854046.2020.1767214] [Medline: 32673163]




XSL•FO
RenderX

33. McAllister TW. Neurobehavioral sequelae of traumatic brain injury: evaluation and management. World Psychiatry 2008 Feb;7(1):3-10 [FREE Full text] [doi: 10.1002/j.2051-5545.2008.tb00139.x] [Medline: 18458777]
34. Geller S. Cultivating online therapeutic presence: strengthening therapeutic relationships in teletherapy sessions. Counselling Psychology Quarterly 2020 Jul 26:1-17. [doi: 10.1080/09515070.2020.1787348]
35. Silverberg ND, Martin P, Panenka WJ. Headache Trigger Sensitivity and Avoidance after Mild Traumatic Brain Injury. J Neurotrauma 2019 May 15;36(10):1544-1550. [doi: 10.1089/neu.2018.6025] [Medline: 30489228]
36. Scharff A, Breiner CE, Ueno LF, Underwood SB, Merritt EC, Welch LM, et al. Shifting a training clinic to teletherapy during the COVID-19 pandemic: a trainee perspective. Counselling Psychology Quarterly 2020 Jun 28:1-11. [doi: 10.1080/09515070.2020.1786668]
37. Hoffman JM, Ehde DM, Dikmen S, Dillworth T, Gertz K, Kincaid C, et al. Telephone-delivered cognitive behavioral therapy for veterans with chronic pain following traumatic brain injury: Rationale and study protocol for a randomized controlled trial study. Contemp Clin Trials 2019 Jan;76:112-119. [doi: 10.1016/j.cct.2018.12.004] [Medline: 30553077]
38. Williams R, Farquharson L, Palmer L, Bassett P, Clarke J, Clark DM, et al. Patient preference in psychological treatment and associations with self-reported outcome: national cross-sectional survey in England and Wales. BMC Psychiatry 2016 Jan 15;16(1):4 [FREE Full text] [doi: 10.1186/s12888-015-0702-8] [Medline: 26768890]
39. Kooistra LC, Wiersma JE, Ruwaard J, Neijenhuijs K, Lokkerbol J, van Oppen P, et al. Cost and Effectiveness of Blended Versus Standard Cognitive Behavioral Therapy for Outpatients With Depression in Routine Specialized Mental Health Care: Pilot Randomized Controlled Trial. J Med Internet Res 2019 Oct 29;21(10):e14261 [FREE Full text] [doi: 10.2196/14261] [Medline: 31663855]
40. Asemota AO, George BP, Cumpsty-Fowler CJ, Haider AH, Schneider EB. Race and insurance disparities in discharge to rehabilitation for patients with traumatic brain injury. J Neurotrauma 2013 Dec 15;30(24):2057-2065 [FREE Full text] [doi: 10.1089/neu.2013.3091] [Medline: 23972035]
41. Backhaus A, Agha Z, Maglione ML, Repp A, Ross B, Zuest D, et al. Videoconferencing psychotherapy: a systematic review. Psychol Serv 2012 May;9(2):111-131. [doi: 10.1037/a0027924] [Medline: 22662727]

## Abbreviations

**BDI-II:** Beck Depression Inventory-II
**CBT:** cognitive behavioral therapy
**CBT-TBI:** cognitive behavioral therapy for depression in individuals with traumatic brain injury
**MDD:** major depressive disorder
**NIH:** National Institutes of Health
**RCT:** randomized controlled trial
**STTS-R:** Satisfaction with Therapy and Therapist Scale-Revised
**TBI:** traumatic brain injury

*Edited by G Eysenbach; submitted 12.03.21; peer-reviewed by R van Eijk, B Smith; comments to author 12.07.21; revised version received 18.09.21; accepted 05.10.21; published 01.12.21.*

*Please cite as:*
*Fisher LB, Tuchman S, Curreri AJ, Markgraf M, Nyer MB, Cassano P, Iverson GL, Fava M, Zafonte RD, Pedrelli P*
*Transitioning From In-Person to Remote Clinical Research on Depression and Traumatic Brain Injury During the COVID-19 Pandemic: Study Modifications and Preliminary Feasibility From a Randomized Controlled Pilot Study*
*JMIR Form Res 2021;5(12):e28734*
*URL: https://formative.jmir.org/2021/12/e28734*
*doi:10.2196/28734*
*PMID:34662285*

©Lauren B Fisher, Sylvie Tuchman, Andrew J Curreri, Maggie Markgraf, Maren B Nyer, Paolo Cassano, Grant L Iverson, Maurizio Fava, Ross D Zafonte, Paola Pedrelli. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Depression, Anxiety, and Daily Activity Among Adolescents Before and During the COVID-19 Pandemic: Cross-sectional Survey Study

Anna Jolliff[1], MS; Qianqian Zhao[1], MS; Jens Eickhoff[1], PhD; Megan Moreno[1], MD, MSEd, MPH

Department of Pediatrics, School of Medicine and Public Health, University of Wisconsin - Madison, Madison, WI, United States

**Corresponding Author:**
Anna Jolliff, MS
Department of Pediatrics
School of Medicine and Public Health
University of Wisconsin - Madison
2870 University Avenue
Suite 200
Madison, WI, 53705
United States
Phone: 1 5035539361
Email: ajolliff@wisc.edu

## Abstract

**Background:** The COVID-19 pandemic has resulted in significant changes to adolescents' daily lives and, potentially, to their mental health. The pandemic has also disproportionately affected historically marginalized and at-risk communities, including people of color, socioeconomically disadvantaged people, people identifying as female, and youth.

**Objective:** This study aimed to understand differences in depression and anxiety among 2 groups of adolescents in the United States before and during the COVID-19 pandemic, and to examine demographic and daily activity variables associated with depression and anxiety.

**Methods:** Online surveys were distributed in 2019 and 2020. Demographic questions were asked at the time of enrollment, and included participants' age, gender, race and ethnicity, and socioeconomic status (SES). The 8-item Patient Health Questionnaire was used to assess symptoms of depression, and the 7-item Generalized Anxiety Disorder scale was used to assess symptoms of anxiety. A total of 4 pandemic-specific daily activity questions were asked only of the pandemic group. Analyses of covariance compared depression and anxiety between prepandemic and pandemic groups. Demographic and lifestyle variables were included as covariates.

**Results:** The sample comprised a total of 234 adolescents, with 100 participants in the prepandemic group and 134 participants in the pandemic group. Within the pandemic group, 94% (n=126) of adolescents reported being out of school due to the pandemic, and another 85.8% (n=115) and 57.1% (n=76) were prevented from extracurricular activities and exercise, respectively. Higher depression was seen in the pandemic group, with a least-squares adjusted mean of 7.62 (SD 1.36) compared to 6.28 (SD 1.42) in the prepandemic group, although the difference was not significant ($P$=.08). There was no significant difference in anxiety scores between the 2 groups (least-squares adjusted means 5.52, SD 1.30 vs 5.01, SD 1.36; $P$=.48). Within the pandemic group, lower SES was predictive of anxiety, such that those in the pandemic group of lower SES were more anxious than their higher-SES peers (least-squares adjusted means 11.17, SD 2.34 vs 8.66, SD 2.16; $P$=.02). Within the pandemic group, being out of work or school and not partaking in extracurricular activities or exercise due to the pandemic were not associated with higher depression or anxiety scores.

**Conclusions:** In this study, neither being in the pandemic group nor experiencing changes in daily activity due to the pandemic was associated with higher depression or anxiety. However, we found that adolescents from lower SES backgrounds experienced significantly more anxiety during the pandemic than their more privileged peers. Both instrumental and mental health interventions for low-income adolescents are imperative.

***(JMIR Form Res 2021;5(12):e30702)*** doi:10.2196/30702



XSL•FO
RenderX

**KEYWORDS**

COVID-19; pandemic; adolescent; depression; anxiety; socioeconomic status; survey; mental health

## Introduction

### Background

COVID-19 resulted in significant changes to adolescents' daily lives [1]. This pandemic led to a prolonged period of statewide school closures [2,3], social isolation, unemployment [4], millions of infections, and hundreds of thousands of deaths [5]. The pandemic also disproportionately affected historically marginalized and at-risk communities, including people of color, socioeconomically disadvantaged people, people identifying as female, and youth [6,7]. However, it is unclear whether the early days of the pandemic were associated with appreciable changes in adolescents' mental health in the United States.

Early international research suggested that the COVID-19 pandemic was associated with elevated rates of depression and anxiety in adolescents [8]. Research performed in Spain and Italy found increases in anxiety, irritability, and restlessness among adolescents [9], while Chinese adolescents reported increases in depression and anxiety during the pandemic [8,10]. Research in Germany suggests that increases in depression, anxiety, and distress were especially pronounced among those with pre-existing mental illnesses [11]. However, much of the research on the pandemic thus far has focused on either adults or adolescents outside the United States [9]. Limited research on US adolescents during COVID-19 demonstrates increases in depression [12,13] and anxiety [13,14]. A systematic review published in 2021 [15] demonstrated that research on US adolescents was still in a nascent stage; out of 16 international studies on the psychological impacts of COVID-19 on children, only 2 focused on US participants [16,17]. This is an important gap, as the experience of American adolescents has been somewhat unique from that of adolescents in other countries; safety measures were implemented more slowly and with greater variability in the United States, and remained in effect longer than anticipated [18].

Adolescents did not all fare equally well during the COVID-19 pandemic. Existing research suggests that, during COVID-19, adolescent females scored higher in depression and anxiety than adolescent males [8,14,15,19-21]. However, because most of these studies are cross-sectional, it is not clear whether COVID-19 disproportionately affected female over male mental health, or whether the greater anxiety and depression among female adolescents simply reflects pre–COVID-19 differences [22]. Indeed, 1 longitudinal study found that although adolescent females reported greater depression than males during the pandemic, this was proportionate to their pre–COVID-19 differences in depression [20].

Existing research shows that people of lower socioeconomic status (SES) had worse mental health outcomes during COVID-19. Research shows that US adults with a lower income and less than $5000 in savings experienced greater depression during the pandemic [12]. A study of US adolescents showed that residing in a county with higher poverty levels during COVID-19 was associated with greater depression and anxiety [21]. These negative mental health consequences may be due to the disproportionately negative impact COVID-19 had on people of lower SES. This group experienced a higher incidence of COVID-19 [23] due to limited access to health information, lack of protective personal equipment, and reduced freedom to socially distance [24]. Once infected, lower SES also predicted invasive mechanical ventilation, intensive care unit admission, and mortality [23,25]. The reality of greater risk among those with lower SES may explain greater depression and anxiety in this group; however, it is unclear whether these differences would be perceptible in the earliest days of the pandemic in a sample of adolescents of lower SES.

Research shows that, during COVID-19, being a member of a historically marginalized racial or ethnic group was associated with suicidality [26] and depression [12]. The observed decline in mental health may have been due to the disproportionately negative impact of COVID-19 on racial and ethnic minority communities. Members of racial and ethnic minority groups are more likely to face structural inequities, including inadequate access to health care and overcrowded schools and neighborhoods, both of which contribute to COVID-19 transmission and mortality [25,27]. Although racial and ethnic minority status is closely related to SES in the United States [25], Kim and colleagues [28] attempted to disaggregate the effects of minority status from the effects of SES on COVID-19 outcomes. These authors found that racial minority status, independent of SES, predicted COVID-19 death rate [28]. This may be attributable to the intra- and intergenerational effects of systemic racism on African American individuals' health and quality of health care.

Daily activities such as part-time work, school, extracurricular activities, and exercise are a source of social support for many adolescents. The sudden removal of these activities —and therefore, of social support—may have resulted in depression and anxiety among US adolescents. Research on Canadian adolescents has shown that having schoolwork to do during quarantine was associated with less depression, while exercise was associated with less loneliness [29]. Research on adolescent athletes shows that athletes involved in team sports experienced more depression during COVID-19 than their peers involved in individual sports, suggesting that the sudden absence of the social component may have contributed to poor mental health [21]. In 2020, Loades and colleagues [30] conducted a rapid systematic review of pre–COVID-19 studies in an attempt to predict the likely effects of social isolation and loneliness on the mental health of adolescents. Based on their sample of articles, these researchers predicted that social isolation due to COVID-19 would result in increased depression among previously healthy adolescents.

### This Study

In summary, more research is needed to determine whether American adolescents experienced greater anxiety and depression in the early days of the COVID-19 pandemic; whether this effect was more pronounced among female,



XSL•FO
RenderX

socioeconomically disadvantaged, or racial and ethnic minority adolescents; and whether changes to daily activity were associated with negative mental health. This study contrasts the mental health of US adolescents surveyed prior to the pandemic with that of adolescents surveyed 2 weeks after many states had enacted statewide school closures [2]. Our hypotheses were as follows:

1. Participants surveyed during the pandemic will report greater depression and anxiety than participants surveyed prior to the pandemic.
2. Adolescents surveyed during the pandemic who identify as (A) female, (B) having lower SES, or (C) a racial or ethnic minority will report greater depression and anxiety than their peers surveyed during the pandemic.
3. Adolescents who were prevented from attending (A) work, (B) school, (C) extracurricular activities, or (D) exercise due to the pandemic will report greater depression and anxiety than their peers who were not prevented from carrying out these activities.

It is vital to test these hypotheses as a key step to understanding and designing interventions for adolescents experiencing negative mental health outcomes during and in the aftermath of the COVID-19 pandemic.

# Methods

## Overview

This 2-group cross-sectional study was a subproject of a larger study. All surveys were taken online. The first group of participants completed the surveys between October 2019 and February 2020, while the second group of responses were received from March 31 to April 3, 2020. This study was approved by the (blinded for review) Institutional Review Board.

## Recruitment

### *Participants*

Eligible participants were aged 13-17 years, lived in the United States, and spoke English. The aims of the larger project included evaluating different recruitment approaches for participation in ecological momentary assessment research. Thus, potential subjects for this subproject were recruited using 2 recruitment methods, including Qualtrics and Facebook advertisements. The prepandemic group included participants recruited via Qualtrics and Facebook, while the pandemic group included participants recruited via Facebook only.

### *Facebook Recruitment*

Participants were recruited via paid advertisements posted on Facebook. Advertisements were targeted at the parents of teenagers aged 13-17 years. Parents who clicked on the advertisement were invited to read the informed consent document. Those who provided consent were invited to pass the device to their child, who would then provide assent and complete the eligibility screening. Parents whose children were not with them to provide assent were prompted to enter their child's phone number; the child would then receive a link to read the assent document and complete the eligibility screening.

### *Qualtrics Recruitment*

Qualtrics is a service that partners with survey research companies to recruit research participants meeting specified inclusion criteria [31]. Qualtrics samples have been shown to be politically and demographically representative of the US population and can replicate known population-level effects [32,33]. Qualtrics sends survey invitations to eligible participants, and compensates participants with rewards, including gift cards, retail points, and airline miles. In this study, Qualtrics sent invitations to the parents of teenagers aged 13-17 years. Only parents who indicated that their children were currently with them and available to provide assent were allowed to complete the consent document. After providing consent, parents were prompted to pass the device to their child so that their child could provide assent and complete the eligibility screening.

## Survey Measures

### *Demographic Information*

Demographic questions were asked at the time of enrollment, and included participants' age, gender, race and ethnicity, and SES. For age, respondents could select a number between 13 and 17. For gender, response options were "female," "male," "nonbinary gender," "female to male transgender, "male to female transgender," and "other," with the option to write in a response. For race and ethnicity, participants could select 1 or more of the following options: "White," "Black/African American, "Asian or Pacific Islander," "American Indian or Alaska Native, "Hispanic/Latino," and "other," with the option to write in a response. To assess SES, participants were asked whether they or their sibling were on free or reduced lunch, and response options were "yes," "no," and "I don't know."

### *Mental Health Outcomes*

#### Depression

The 8-item Patient Health Questionnaire (PHQ-8) was used to assess symptoms of depression [34]. Participants were asked to rate the extent to which they had experienced certain symptoms, from "not at all" to "nearly every day." Example symptoms include "poor appetite or overeating," "feeling tired or having little energy," and "feeling down, depressed, or hopeless." Cutoff scores of 5, 10, and 15 were used to indicate mild, moderate, and severe depression, respectively. Internal consistency for the PHQ-8 is excellent (Cronbach $\alpha$=.89) [35,36].

#### Anxiety

The 7-item Generalized Anxiety Disorder (GAD-7) scale was used to assess symptoms of anxiety [37]. Similar to the PHQ-8, participants were asked to rate the extent to which they had experienced certain symptoms, from "not at all" to "nearly every day." Example symptoms include "not being able to stop or control worrying" and "feeling afraid as if something awful might happen." As with the PHQ-8, cutoff scores of 5, 10, and 15 were used to indicate mild, moderate, and severe anxiety, respectively [37]. Internal consistency for the GAD-7 is excellent (Cronbach $\alpha$=.92) [37].



XSL•FO
RenderX

### Changes in Daily Activity Due to the Pandemic

A total of 4 pandemic-specific daily activity questions were asked only of the pandemic group. Participants were asked to indicate whether, at present, they were prevented from (1) work, (2) school, (3) extracurricular activities, or (4) exercise due to the pandemic. These were categorical variables, and for each question participants could select 1 of 4 responses: "yes," "no," "does not apply," or "other," with the option to write in a response.

## Procedure

All participants received a link by SMS text messaging to complete the survey. Participants had 3 days after receiving the text message to complete the survey, and participant responses were retained if they completed more than 50% of it. Participants recruited through Qualtrics were compensated through Qualtrics, and those recruited through Facebook were compensated for participation with a check sent to their mailing address.

## Analysis

### *Demographic Variables*

Descriptive statistics were used to summarize participants' age, race and ethnicity, and gender, as well as the percentage of participants who were out of work, school, extracurricular activities, or exercise due to the pandemic. Age was treated as a continuous variable. Given the small number of participants endorsing certain races and ethnicities (eg, Hispanic/Latino, Asian, Pacific Islander), respondents were categorized as non-Hispanic White, non-Hispanic Black, multirace, or other. Due to the small number of participants endorsing genders other than male or female, respondents were categorized as "male," "female," or "other." Participants who indicated that they did not know if they or their siblings were on free or reduced lunch were categorized as not being on free or reduced lunch. Most participants reported being out of school due to the pandemic; thus, the answers "no," "does not apply," and "other" were combined in the multivariate analyses. To evaluate demographic differences between the prepandemic and pandemic groups, the Fisher exact test was used for categorical variables, a 2-sample *t* test was used for normally distributed continuous variables, and a Wilcoxon rank-sum nonparametric test was used for nonnormally distributed continuous variables.

### *Main Analyses*

Statistical analyses were conducted using SAS software, version 9.4 (SAS Institute Inc) [38]. All reported *P* values were 2-sided, and $P<.05$ was used to define statistical significance. Hypothesis 1 predicted that being in the pandemic group would be associated with significantly higher depression and anxiety scores compared to the prepandemic group. Analysis of covariance (ANCOVA) was used to test this hypothesis. Anxiety and depression were entered as continuous dependent variables, rather than categorical ones, in order to maximize sensitivity. Group was entered as an independent variable, and demographic variables (age, race, gender, SES) were entered as covariates.

Hypothesis 2 stated that, within the pandemic group, participants identifying as female, as having a lower SES, or as a racial or ethnic minority would be associated with higher depression and anxiety scores. Hypothesis 3 stated that, within the pandemic group, being out of work, school, extracurricular activities, or exercise due to the pandemic would be associated with higher depression and anxiety scores. ANCOVA was again used to test these hypotheses. Anxiety and depression were entered as continuous dependent variables. Work, school, extracurricular activities, exercise, and demographic variables were entered as covariates.

We conducted a sensitivity analysis by excluding all Qualtrics respondents and rerunning the analyses. Furthermore, we evaluated interaction effects between the groups (pre–COVID-19 or COVID-19) and each demographic variable (age, gender, race and ethnicity, and SES) to examine whether further subgroup analysis by demographic characteristics was required.

# Results

## Participants

There were 100 participants in the prepandemic group, and 134 participants in the pandemic group, for a combined total of 234 participants. Demographic differences between the 2 groups are recorded in Table 1. The samples showed a similar distribution of gender, although statistically significant differences were noted in race, age, and the number of participants on free or reduced lunch. Within the prepandemic group, the mean PHQ-8 and GAD-7 scores were 5.27 (SD 5.06) and 4.67 (SD 5.22), respectively. In the pandemic group, the mean PHQ-8 and GAD-7 scores were 6.81 (SD 5.67) and 5.37 (SD 5.18), respectively. Regarding changes in daily activity during the pandemic, 94% (126/134) of participants reported that they were prevented from attending school during the pandemic, and 85.8% (115/134) were prevented from extracurricular activities. The percentage of participants out of work, school, extracurricular activities, or exercise due to the pandemic is shown in Table 2.



XSL•FO
RenderX

**Table 1.** Demographic information for prepandemic and pandemic samples (N=234).

| Characteristic | Prepandemic (n=100) | Pandemic (n=134) | *P* value |
|---|---|---|---|
| Age (years), mean (SD) | 14.7 (1.3) | 15.2 (1.4) | .01[a] |
| **Gender, n (%)** | | | .49[b] |
| Female | 66 (66) | 92 (68.7) | |
| Male | 34 (34) | 40 (29.9) | |
| Other | 0 (0) | 2 (0.01) | |
| **Race, n (%)** | | | <.001[b] |
| White, non-Hispanic | 56 (56) | 103 (76.9) | |
| Black, non-Hispanic | 18 (18) | 8 (6) | |
| Multirace | 11 (11) | 12 (9) | |
| Other | 15 (15) | 11 (8.2) | |
| **Qualified for free or reduced lunch, n (%)** | | | <.001[b] |
| Yes | 51 (51) | 41 (30.6) | |
| No/don't know | 49 (49) | 93 (69.4) | |

[a]*P* value from 2-sample *t* test or Wilcoxon nonparametric test.
[b]*P* value from the Fisher exact test.

**Table 2.** Participants out of work, school, extracurricular activities, or exercise due to the COVID-19 pandemic 2 weeks after nationwide school closures (n=134).

| Activity | Participants |
|---|---|
| **Work, n (%)** | |
| Yes | 31 (23.1) |
| No | 5 (3.7) |
| Does not apply | 98 (73.1) |
| Other | 0 (0) |
| **School, n (%)** | |
| Yes | 126 (94.0) |
| No | 4 (3.0) |
| Does not apply | 2 (1.5) |
| Other | 2 (1.5) |
| **Extracurricular activities, n (%)** | |
| Yes | 115 (85.8) |
| No | 4 (3.0) |
| Does not apply | 15 (11.2) |
| Other | 0 (0) |
| **Exercise, n (%)** | |
| Yes | 76 (57.1) |
| No | 28 (21.1) |
| Does not apply | 20 (15.0) |
| Other | 9 (6.8) |



XSL•FO
RenderX

### Hypothesis Testing

Hypothesis 1 stated that being in the pandemic group would be associated with significantly higher depression and anxiety scores. In the adjusted model for depression, the pandemic group exhibited higher depression, with a least-squares adjusted mean of 7.62 (SD 1.36) compared to 6.28 (SD 1.42) in the prepandemic group; however, this difference was statistically insignificant ($P$=.08). Across both groups, female participants had greater depression scores than males (least-squares adjusted means 6.45, SD 0.55 vs 4.70, SD 0.69; $P$=.02). For anxiety, there was no significant difference in anxiety scores between the pandemic and prepandemic groups (least-squares adjusted means 5.52, SD 1.30 vs 5.01, SD 1.36; $P$=.48). Again, across both groups, females reported greater anxiety than males (least-squares adjusted means 5.59, SD 0.53 vs 3.58, SD 0.66; $P$=.01). No other demographic variables were significant predictors of depression or anxiety.

Hypothesis 2 stated that, within the pandemic group, identifying as female, as having a lower SES, or as a racial or ethnic minority would be associated with greater depression and anxiety. In the adjusted model for depression, none of these variables were significant predictors of depression. In the adjusted model for anxiety, being eligible for free or reduced lunch was associated with greater anxiety, such that those in the pandemic group who qualified for free or reduced lunch were more anxious than their peers who did not (least-squares adjusted means 11.17, SD 2.34 vs 8.66, SD 2.16; $P$=.02).

Hypothesis 3 stated that, within the pandemic group, being out of work, school, extracurricular activities, or exercise due to the pandemic would be associated with higher depression and anxiety scores. No significant associations were detected.

When we conducted a sensitivity analysis by excluding the Qualtrics participants and rerunning the analyses, the results did not change. We also did not detect any significant interaction effect between the groups (pre–COVID-19 or COVID-19) and the demographic characteristics. Therefore, no additional subgroup analyses by demographic characteristics were conducted.

## Discussion

### Principal Findings

This study examined differences in mental health between adolescents surveyed prior to the COVID-19 pandemic and those surveyed approximately 2 weeks after many states had enacted statewide school closures. Contrary to our hypothesis, adolescents in the pandemic group did not score significantly higher in depression and anxiety after adjusting for covariates. Within the pandemic group, lower SES was associated with higher anxiety. The majority of adolescents were prevented from attending school or participating in extracurricular activities and exercise due to the pandemic, and these changes in daily activity were not associated with depression or anxiety. These findings have important implications for future research on the adolescent experience of the COVID-19 pandemic.

First, we expected to find higher depression and anxiety scores in the pandemic group. This potential finding would have been consistent with existing international literature on COVID-19, which has shown that adolescents and young adults experienced elevated depression and anxiety due to the pandemic [8-10]. In our study, after controlling for the effects of demographic variables, being in the pandemic group was not associated with higher depression or anxiety. It may be that, after only 2 weeks in quarantine, American adolescents did not experience the pandemic as depressing or anxiety-inducing. They may have thought that the restrictions would quickly be lifted or, like many, may not have foreseen the gravity of the virus. Research completed several months into quarantine could yield different results. However, in both the prepandemic and pandemic groups, being female was associated with greater PHQ-8 and GAD-7 scores. This is consistent with existing literature, showing that adolescent females are more at risk for depression and anxiety [22].

In the group of participants surveyed during the pandemic, neither gender nor race was associated with depression or anxiety, which is inconsistent with previously cited research on gender [8,20,21] and race [25-28] during COVID-19. However, SES was associated with anxiety, such that participants of lower SES reported greater anxiety. This is consistent with existing research, which showed that SES was a risk factor for worse mental health outcomes during the pandemic [12,21,39]. It may be that COVID-19 magnified disparities that were not statistically significant in the prepandemic sample. Families of lower SES encountered more COVID-19 infections and fatalities, which may directly challenge the mental health of both infected people and their loved ones [23,25,40]. Underresourced adolescents, in particular, often rely on school for affordable food and school-based health care [41], and may have experienced greater anxiety when those resources were no longer available. In the wake of COVID-19, interventions should focus on adolescents of lower SES as their well-being is uniquely at risk due to the pandemic.

Our results showed that 2 weeks after many school closures, 94% of participants were prevented from going to school due to the pandemic. The majority of participants were prevented from partaking in extracurricular activities or exercise due to the pandemic, perhaps because these opportunities are tied to school for many adolescents. However, these disruptions to daily life within the pandemic group were not associated with higher depression and anxiety scores. This may have occurred for multiple reasons. First, only 2 weeks into quarantine, the absence of daily activities may have felt like a welcome reprieve, especially for adolescents with social anxiety [42], as well as students who experience bullying [43] or discrimination [44,45] in the context of their daily activities. Second, adolescents in this study may have experienced more opportunities for prosocial behavior as a result of the pandemic, such as helping, caring for, and comforting friends and family members; prosocial behavior is associated with positive adjustment [46]. Last, adolescents have a demonstrated ability to use technology to navigate developmental and mental health challenges; in our study, adolescents may have found entertainment, support, and distraction through digital interaction, such that any effects of COVID-19 on mental health were mitigated [47].



XSL•FO
RenderX

## Limitations

This study was not without limitations. First, the prepandemic and pandemic groups were significantly different in terms of age, race and ethnicity, and qualifying for free or reduced lunch. Additionally, prepandemic data were collected in the fall and winter months, while pandemic data was collected in the spring. This may have obscured differences caused by the pandemic. However, with the exception of season, ANCOVA should have controlled for all group differences. Second, the majority of participants did not work, making it impossible to ascertain the effects of the pandemic on adolescents who lost employment. Third, there may be constructs better suited than "anxiety" and "depression" for understanding adolescents' experience of the pandemic. Negative experiences such as fear, boredom, or confusion may be more appropriate, as might positive experiences such as feeling less social or academic pressure, or feeling more supportive and supported online and offline. Further, the 4 pandemic-specific daily activity questions were not based on any validated scales, so these results should be interpreted with caution. Last, because data were collected in early April, findings may not be generalizable to the later periods of the pandemic.

## Conclusion

This research suggested that, at least in the early days of the pandemic, adolescents had not developed significantly greater anxiety or depression at the population level. Further, 2 weeks after many school closures, reductions in daily activities were not associated with anxiety and depression. However, we found that adolescents from lower socioeconomic backgrounds experienced significantly more anxiety during the pandemic than their peers of higher SES. Both instrumental and mental health interventions for adolescents of lower SES are imperative. More research is needed to understand the trajectory of adolescents' mental health experiences during the COVID-19 pandemic, as well as the long-term impact of the pandemic on adolescent mental health.

## Acknowledgments

This work was supported by a collaborative research agreement with Facebook. We would like to thank the adolescents who participated in this study.

## Conflicts of Interest

None declared.

## References

1. Stress in the time of COVID-19. American Psychological Association. 2020. URL: https://tinyurl.com/yycjp8ra [accessed 2021-10-12]
2. School responses to the coronavirus (COVID-19) pandemic during the 2019-2020 academic year. Ballotpedia. 2020. URL: https://ballotpedia.org/School_responses_to_the_coronavirus_(COVID-19)_pandemic_during_the_2019-2020_academic_year [accessed 2021-10-13]
3. Christakis DA, Van Cleve W, Zimmerman FJ. Estimation of US children's educational attainment and years of life lost associated with primary school closures during the coronavirus disease 2019 pandemic. JAMA Netw Open 2020 Nov 02;3(11):e2028786 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.28786] [Medline: 33180132]
4. Ember S. Unemployment Claims Rise Sharply, Showing New Economic Pain. The New York Times. 2021 Jan 14. URL: https://www.nytimes.com/2021/01/14/business/economy/unemployment-claims.html [accessed 2021-10-13]
5. CDC COVID Data Tracker. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker/#datatracker-home [accessed 2021-10-13]
6. Clark E, Fredricks K, Woc-Colburn L, Bottazzi ME, Weatherhead J. Disproportionate impact of the COVID-19 pandemic on immigrant communities in the United States. PLoS Negl Trop Dis 2020 Jul 13;14(7):e0008484 [FREE Full text] [doi: 10.1371/journal.pntd.0008484] [Medline: 32658925]
7. Hatef E, Chang H, Kitchen C, Weiner JP, Kharrazi H. Assessing the impact of neighborhood socioeconomic characteristics on COVID-19 prevalence across seven states in the United States. Front Public Health 2020 Sep 22;8:571808 [FREE Full text] [doi: 10.3389/fpubh.2020.571808] [Medline: 33072710]
8. Duan L, Shao X, Wang Y, Huang Y, Miao J, Yang X, et al. An investigation of mental health status of children and adolescents in China during the outbreak of COVID-19. J Affect Disord 2020 Oct 01;275:112-118 [FREE Full text] [doi: 10.1016/j.jad.2020.06.029] [Medline: 32658812]
9. Imran N, Aamer I, Sharif MI, Bodla ZH, Naveed S. Psychological burden of quarantine in children and adolescents: A rapid systematic review and proposed solutions. Pak J Med Sci 2020 Jun 24;36(5):1106-1116 [FREE Full text] [doi: 10.12669/pjms.36.5.3088] [Medline: 32704298]
10. Chen F, Zheng D, Liu J, Gong Y, Guan Z, Lou D. Depression and anxiety among adolescents during COVID-19: A cross-sectional study. Brain Behav Immun 2020 Aug;88:36-38 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.061] [Medline: 32464156]



XSL•FO
RenderX

11. Bäuerle A, Steinbach J, Schweda A, Beckord J, Hetkamp M, Weismüller B, et al. Mental health burden of the COVID-19 outbreak in Germany: predictors of mental health impairment. J Prim Care Community Health 2020 Aug 29;11:2150132720953682-2150132720953688 [FREE Full text] [doi: 10.1177/2150132720953682] [Medline: 32865107]
12. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
13. Liu CH, Zhang E, Wong GTF, Hyun S, Hahm. Factors associated with depression, anxiety, and PTSD symptomatology during the COVID-19 pandemic: Clinical implications for U.S. young adult mental health. Psychiatry Res 2020 Aug;290:113172 [FREE Full text] [doi: 10.1016/j.psychres.2020.113172] [Medline: 32512357]
14. Hawes MT, Szenczy AK, Klein DN, Hajcak G, Nelson BD. Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic. Psychol. Med 2021 Jan 13:1-9. [doi: 10.1017/s0033291720005358]
15. Jones EAK, Mitra AK, Bhuiyan AR. Impact of COVID-19 on mental health in adolescents: a systematic review. Int J Environ Res Public Health 2021 Mar 03;18(5):2470 [FREE Full text] [doi: 10.3390/ijerph18052470] [Medline: 33802278]
16. Fish JN, McInroy LB, Paceley MS, Williams ND, Henderson S, Levine DS, et al. "I'm kinda stuck at home with unsupportive parents right now": LGBTQ youths' experiences with COVID-19 and the importance of online support. J Adolesc Health 2020 Sep;67(3):450-452 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.06.002] [Medline: 32591304]
17. Oosterhoff B. Adolescents' motivations to engage in social distancing during the COVID-19 pandemic: associations with mental and social health. PsyArXiv Preprint posted online Apr 6, 2020. [doi: 10.31234/osf.io/jd2kq]
18. Khanna R, Cicinelli M, Gilbert S, Honavar S, Murthy GV. COVID-19 pandemic: Lessons learned and future directions. Indian J Ophthalmol 2020;68(5):703. [doi: 10.4103/ijo.ijo_843_20]
19. Alonzi S, La Torre A, Silverstein MW. The psychological impact of preexisting mental and physical health conditions during the COVID-19 pandemic. Psychol Trauma 2020 Aug;12(S1):S236-S238. [doi: 10.1037/tra0000840] [Medline: 32525380]
20. Wright N, Hill J, Sharp H, Pickles A. Interplay between long-term vulnerability and new risk: Young adolescent and maternal mental health immediately before and during the COVID-19 pandemic. JCPP Adv 2021 Jan 18;1(1):e12008 [FREE Full text] [doi: 10.1111/jcv2.12008] [Medline: 34485987]
21. McGuine T, Biese KM, Petrovska L, Hetzel SJ, Reardon C, Kliethermes S, et al. Mental health, physical activity, and quality of life of US adolescent athletes during COVID-19-related school closures and sport cancellations: a study of 13.000 athletes. J Athl Train 2021 Jan 01;56(1):11-19 [FREE Full text] [doi: 10.4085/1062-6050-0478.20] [Medline: 33290516]
22. Merikangas KR, He J, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2010 Oct;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
23. Clouston SA, Natale G, Link BG. Socioeconomic inequalities in the spread of coronavirus-19 in the United States: A examination of the emergence of social inequalities. Soc Sci Med 2021 Jan;268:113554 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113554] [Medline: 33308911]
24. Consolazio D, Murtas R, Tunesi S, Gervasi F, Benassi D, Russo AG. Assessing the impact of individual characteristics and neighborhood socioeconomic status during the COVID-19 pandemic in the provinces of Milan and Lodi. Int J Health Serv 2021 Jul 02;51(3):311-324. [doi: 10.1177/0020731421994842] [Medline: 33650453]
25. Quan D, Luna Wong L, Shallal A, Madan R, Hamdan A, Ahdi H, et al. Impact of race and socioeconomic status on outcomes in patients hospitalized with COVID-19. J Gen Intern Med 2021 May 27;36(5):1302-1309 [FREE Full text] [doi: 10.1007/s11606-020-06527-1] [Medline: 33506402]
26. Czeisler, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
27. Rothe EM, Fortuna LR, Tobon AL, Postlethwaite A, Sanchez-Lacay JA, Anglero YL. Structural inequities and the impact of COVID-19 on Latinx children: implications for child and adolescent mental health practice. J Am Acad Child Adolesc Psychiatry 2021 Jun;60(6):669-671. [doi: 10.1016/j.jaac.2021.02.013] [Medline: 33662497]
28. Kim SJ, Bostwick W. Social vulnerability and racial inequality in COVID-19 deaths in Chicago. Health Educ Behav 2020 Aug 21;47(4):509-513 [FREE Full text] [doi: 10.1177/1090198120929677] [Medline: 32436405]
29. Ellis WE, Dumas TM, Forbes LM. Physically isolated but socially connected: Psychological adjustment and stress among adolescents during the initial COVID-19 crisis. Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement 2020 Jul;52(3):177-187. [doi: 10.1037/cbs0000215]
30. Loades ME, Chatburn E, Higson-Sweeney N, Reynolds S, Shafran R, Brigden A, et al. Rapid systematic review: the impact of social isolation and loneliness on the mental health of children and adolescents in the context of COVID-19. J Am Acad Child Adolesc Psychiatry 2020 Nov;59(11):1218-1239.e3 [FREE Full text] [doi: 10.1016/j.jaac.2020.05.009] [Medline: 32504808]
31. Snow J, Mann M, Page T. Qualtrics Survey Software: Handbook for Research Professionals. Provo, Utah: Qualtrics Labs, Inc; 2012.




XSL•FO
RenderX

32. Boas TC, Christenson DP, Glick DM. Recruiting large online samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics. PSRM 2018 Aug 08;8(2):232-250. [doi: 10.1017/psrm.2018.28]
33. Kraiger K, McGonagle AK, Sanchez DR. What's in a sample? Comparison of effect size replication and response quality across student, MTurk, and Qualtrics samples. 2020 Presented at: 11th Conference on Organizational Psychology: People and Risks; April 24, 2020; Saratov State University URL: https://www.sgu.ru/sites/default/files/samples_paper.pdf
34. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
35. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
36. Shin C, Lee S, Han K, Yoon H, Han C. Comparison of the usefulness of the PHQ-8 and PHQ-9 for screening for major depressive disorder: analysis of psychiatric outpatient data. Psychiatry Investig 2019 Apr;16(4):300-305 [FREE Full text] [doi: 10.30773/pi.2019.02.01] [Medline: 31042692]
37. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
38. SAS software, version 9. Cary, NC: SAS Institute; 2002. URL: https://www.sas.com/en_us/home.html [accessed 2021-10-13]
39. Russell MA, Odgers CL. Adolescents' subjective social status predicts day-to-day mental health and future substance use. J Res Adolesc 2020 Feb 02;30 Suppl 2(S2):532-544. [doi: 10.1111/jora.12496] [Medline: 30938467]
40. Mesa Vieira C, Franco OH, Gómez Restrepo C, Abel T. COVID-19: The forgotten priorities of the pandemic. Maturitas 2020 Jun;136:38-41 [FREE Full text] [doi: 10.1016/j.maturitas.2020.04.004] [Medline: 32386664]
41. Mason-Jones AJ, Crisp C, Momberg M, Koech J, De Koker P, Mathews C. A systematic review of the role of school-based healthcare in adolescent sexual, reproductive, and mental health. Syst Rev 2012 Oct 26;1(49):49 [FREE Full text] [doi: 10.1186/2046-4053-1-49] [Medline: 23098138]
42. Morrissette M. School closures and social anxiety during the COVID-19 pandemic. J Am Acad Child Adolesc Psychiatry 2021 Jan;60(1):6-7 [FREE Full text] [doi: 10.1016/j.jaac.2020.08.436] [Medline: 32890669]
43. Chawla N, Sharma P, Sagar R. Psychological impact of COVID-19 on children and adolescents: is there a silver lining? Indian J Pediatr 2021 Jan 14;88(1):91-91 [FREE Full text] [doi: 10.1007/s12098-020-03472-z] [Medline: 32797389]
44. Joseph NM, Viesca KM, Bianco M. Black female adolescents and racism in schools: experiences in a colorblind society. The High School Journal 2016;100(1):4-25. [doi: 10.1353/hsj.2016.0018]
45. Berkel C, Murry VM, Hurt TR, Chen Y, Brody GH, Simons RL, et al. It takes a village: protecting rural African American youth in the context of racism. J Youth Adolesc 2009 Feb 23;38(2):175-188 [FREE Full text] [doi: 10.1007/s10964-008-9346-z] [Medline: 19636716]
46. El Mallah S. Conceptualization and measurement of adolescent prosocial behavior: looking back and moving forward. J Res Adolesc 2020 Jan 18;30 Suppl 1(S1):15-38. [doi: 10.1111/jora.12476] [Medline: 30775824]
47. Ehrenreich SE, Beron KJ, Burnell K, Meter DJ, Underwood MK. How adolescents use text messaging through their high school years. J Res Adolesc 2020 Jun 23;30(2):521-540. [doi: 10.1111/jora.12541] [Medline: 31868974]

## Abbreviations

**ANCOVA:** analysis of covariance
**GAD-7:** 7-item Generalized Anxiety Disorder
**PHQ-8:** 8-item Patient Health Questionnaire
**SES:** socioeconomic status

*Edited by G Eysenbach; submitted 25.05.21; peer-reviewed by FHC Wong, Y Cao; comments to author 21.06.21; revised version received 30.07.21; accepted 27.09.21; published 02.12.21.*

*Please cite as:*
*Jolliff A, Zhao Q, Eickhoff J, Moreno M*
*Depression, Anxiety, and Daily Activity Among Adolescents Before and During the COVID-19 Pandemic: Cross-sectional Survey Study*
*JMIR Form Res 2021;5(12):e30702*
*URL: https://formative.jmir.org/2021/12/e30702*
*doi:10.2196/30702*
*PMID:34609316*

©Anna Jolliff, Qianqian Zhao, Jens Eickhoff, Megan Moreno. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons



XSL•FO
RenderX

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.



XSL•FO
RenderX

Original Paper

# Pandemic Information Dissemination and Its Associations With the Symptoms of Mental Distress During the COVID-19 Pandemic: Cross-sectional Study

Ole Myklebust Amundsen[1]; Asle Hoffart[2,3], ClinPsy, PsyD; Sverre Urnes Johnson[2,3], ClinPsy, PsyD; Omid V Ebrahimi[2,3], ClinPsy

[1]Department of Clinical Psychology, University of Bergen, Bergen, Norway
[2]Modum Bad Psychiatric Hospital, Vikersund, Norway
[3]Department of Psychology, University of Oslo, Oslo, Norway

**Corresponding Author:**
Ole Myklebust Amundsen
Department of Clinical Psychology
University of Bergen
Årstadveien 17
Bergen, 5009
Norway
Phone: 47 45209507
Email: oleamundsen@me.com

## Abstract

**Background:** The 2020-2021 COVID-19 pandemic has added to the mental health strain on individuals and groups across the world in a variety of ways. Viral mitigation protocols and viral spread affect people on all continents every day, but at widely different degrees. To understand more about the mental health consequences of the pandemic, it is important to investigate whether or how people gather pandemic-related information and how obtaining this information differentially affects individuals.

**Objective:** This study aimed to investigate whether and to what extent higher levels of COVID-19–related media consumption across information sources are associated with the symptoms of anxiety, health anxiety, and depression, and whether and to what extent using social media and online interactive platforms versus traditional media platforms is associated with the symptoms of anxiety, health anxiety, and depression. Additionally, we aimed to investigate whether and to what extent avoidance of COVID-19–related information is associated with the aforementioned symptoms.

**Methods:** In a cross-sectional preregistered survey, 4936 participants responded between June 22 and July 13, 2020. Eligible participants were adults currently residing in Norway and were thus subjected to identical viral mitigation protocols. This sample was representative of the Norwegian population after utilizing an iterative raking algorithm to conduct poststratification. As 2 subgroups (transgender and intersex individuals) were too small to be analyzed, the final sample for descriptive statistics and regressions included 4921 participants. Multiple regressions were used to investigate associations between the symptoms of psychopathology and COVID-19–related information dissemination. Part correlations were calculated as measures of the effect size for each predictor variable. Due to the large anticipated sample size, the preregistered criterion for significance was set at $P<.01$.

**Results:** The symptoms of anxiety and health anxiety were significantly associated with obtaining information from newspapers ($P<.001$), social media ($P<.001$), and the broader categories of online interactive ($P<.001$) and traditional media ($P<.001$). The symptoms of depression were significantly associated with obtaining information from newspapers ($P=.003$), social media ($P=.009$), and the broader category of online interactive media ($P<.001$). Additionally, avoidance of COVID-19–related information showed a significant association in all 3 domains of psychopathological symptoms (anxiety and depression, $P<.001$; health anxiety, $P=.007$).

**Conclusions:** This study found significant associations between the symptoms of psychopathology and the use of media for obtaining information related to the COVID-19 pandemic. Significant findings for obtaining information through newspapers, social media, and online interactive media were seen across all 3 measures of psychopathology. Avoidance of COVID-19–related information and associations with the symptoms of psychopathology emerged as core findings, with generally higher effect sizes compared with information attainment.



XSL•FO
RenderX

**Trial Registration:** ClinicalTrials.gov NCT04442360; https://clinicaltrials.gov/ct2/show/NCT04442360

***(JMIR Form Res 2021;5(12):e28239)*** doi:10.2196/28239

**KEYWORDS**

information sources; COVID-19; avoidance; psychopathology

## Introduction

### Background

As of September 2021, the COVID-19 (SARS-CoV-2) pandemic has led to more than 4.5 million deaths worldwide and more than 218 million confirmed cases [1]. While governments strive to provide their inhabitants with up-to-date and accurate information, deliberate and unintentional misinformation has spread across the world [2]. Efficacious information delivery is critical to inform inhabitants of national, regional, and local viral mitigation protocols (VMPs) that are implemented in lieu of a vaccine. Access to mental health services has become an important issue during protracted pandemic restrictions and VMPs [3], and the psychological consequences of the pandemic have been elaborated upon, but its possible long-term consequences are not yet known [4,5].

Imperative pandemic-related information needs to be distributed appropriately to societal members. Concurrently, scientists and government officials have become increasingly concerned with the mental health consequences of VMPs, such as lockdowns, curfew, and more generally forced social isolation, during the COVID-19 pandemic [6]. The scientific community has urged the importance of investigating how high exposure to health messaging disseminated across divergent media platforms is related to mental health. There are indications of increased psychopathological symptoms when attempting to mitigate physical disease and viral spread [6-9].

Although information distribution and access are crucial to inform citizens about changing VMPs, steps to increase quality of life despite restrictions and information on how to live life under lockdown and restriction would be beneficial. However, the world is currently experiencing an overabundance of information, which is referred to as an *infodemic* by the World Health Organization to highlight that there is an ongoing information pandemic parallel to the viral pandemic. Earlier research has shown that overly negative news may lead to aversion and consequently to avoidance of news [10].

In the current COVID-19 pandemic, the level of contagion of many countries has risen sharply over time, or has risen and fallen in waves. Additionally, news has been dominated by high levels of contagion and death, and has been repeated incessantly on national and international news channels. Some research has also shown that individuals have a propensity toward passing along negative news even if the news is exaggerated or there is more positive news that is equally relevant at hand, and the current global infodemic may partly be made up of the spread of such frequently repeated negative news [11].

Studies from previous pandemics (eg, Zika, swine flu, and Ebola) and the current pandemic have revealed that individuals highly engaged in information-seeking behavior tend to become more focused on the threatening aspects of infectious diseases, experiencing increased distress and anxiety [12,13]. This is especially so when individuals perceive themselves as having little control over the threat [13].

Several years of research have given credence to the idea that increased consumption of information from social media in general can be detrimental to psychological health [14-19], and that increased use of such media is associated with the symptoms of both anxiety and depression [17,20-22]. Social media (eg, Twitter, Facebook, TikTok, Snapchat, Instagram, and WeChat), characterized by fast-paced information streams, are often not curated or validated for facts [23] and have in recent years been an arena for the spread of misinformation [24]. In the current pandemic, social media have been used to dispense information more rapidly to the population [25].

Information is also actively spread through other platforms, such as television, newspapers, websites, and online forums. Additionally, individuals are also informed about the pandemic face-to-face via their peers. News distributed on television can be especially volatile. A study found that the median airtime for medical news stories was just 33 seconds, and these stories did not cite the origin of the information they provided and did not convey recommendations to the public [26]. The association of consumption of information obtained via traditional media with mental distress has also been investigated, but the interaction is less clear than it is for social media [27].

Furthermore, the avoidance of news can have direct consequences for individuals and the society at large. In the current infodemic, news avoidance may be a consequence of news overload (ie, unintentional avoidance). During the current pandemic, relevant advice from intergovernmental, national, and regional authorities has been constantly changing, putting a large strain on individuals who want to keep themselves updated on the current recommendations for hygienic behavior and social distancing. An overload of information sources from platforms, such as newspapers and television, may cause those avoiding these platforms to seek pandemic-relevant information from other sources, such as friends and family, which has been found to be related to decreased adherence to social distancing protocols [28]. Thus, understanding and combating avoidance of information, specifically from official sources, is crucial.

Avoiding information can entail not being confronted with uncertainty. In crises, such as the current pandemic, large groups of people are required to live life with a far higher degree of everyday uncertainty than usual [29]. Individuals with high intolerance of uncertainty attempt to reduce this uncertainty through behavioral control, exemplified as checking the internet and other sources for information [29-31]. Furthermore, health anxiety is tied to an increase in the search for health-related information, with psychological distress as a consequence [32].



XSL•FO
RenderX

With an increase in the tendency to search for information, there is also an increase in different media that provide the information, with potentially differential effects.

Thus, it is crucial to investigate how dissemination of divergent sources of information is related to mental health symptoms in the general population during the present COVID-19 pandemic and parallel infodemic.

### Research Questions and Hypotheses

The following research questions and hypotheses were investigated, as presented in the preregistered protocol of the study:

1. Research questions: Is there a differential effect among different information sources on health anxiety, depression, and anxiety? To what extent and how are different information sources related to the symptoms of health anxiety, depression, and anxiety?

2. Hypothesis 1: Media consumption across all information sources will significantly be associated with depression and anxiety symptoms, with increased media consumption in general associated with higher levels of health anxiety, depression, and general anxiety.

3. Hypothesis 2: Using social media and online interactive platforms (ie, forums and blogs) to obtain news about the pandemic in comparison to using traditional media (ie, television, radio, and newspapers) will be associated with higher levels of general anxiety, depression, and health anxiety. Actively staying away from information will further significantly be associated with higher levels of anxiety, depression, and health anxiety.

## Methods

### Overview

This cross-sectional study is part of The Norwegian COVID-19 Mental Health and Adherence Project, and utilizes data collected in the second stage of data collection in this project. This study investigated the association of various sources of information acquisition concerning COVID-19 with psychopathology, specifically the symptoms of anxiety, depression, and health anxiety during the COVID-19 pandemic. Following the guidelines of Strengthening the Reporting of Observational Studies in Epidemiology [33], and the health estimate reporting standards laid out in the GATHER statement [34], this study was designed and preregistered prior to any data collection. The preregistered protocol can be found at ClinicalTrials.gov (NCT04442360). All components of the submitted study adhere to its preregistered protocol.

### Study Design, Participants, Procedure, and Timing

The data for this cross-sectional study were collected via an online survey between June 22 and July 13, 2020, and involved the second stage of data collection for the project. In the first wave of data collection, a survey was distributed on national, regional, and local information platforms (ie, television, radio, and newspapers), in addition to dissemination to a random selection of Norwegian adults through a Facebook Business algorithm. Details regarding procedure and timing can be found elsewhere [9]. In the first stage, data were obtained from 10,061 participants. For the second stage of data collection, all participants who had provided informed consent to participate further were invited to take part. The eligible participants were adults (ie, aged 18 years or above) currently residing in Norway, who were thus subject to identical VMPs. A total of 4936 participants made up the sample for this study. The survey was administered approximately 1 week after major VMPs were lifted in Norway and included a period when the national VMPs and guidelines did not shift. Additionally, no novel information was provided by the Norwegian government regarding social distancing protocols during this period.

### Measurements

#### *Demographic Characteristics*

Information regarding participant age, sex, education level, ethnic background, and regional affiliation level were collected.

#### *Symptoms of Psychopathology*

The Generalized Anxiety Disorder-7 (GAD-7) [35] is a scale for identifying the level of anxiety. It consists of 7 items measuring anxiety on a 4-point Likert scale (range 0-3), with the total score ranging from 0 to 21. For GAD-7, internal consistency was good, with a Cronbach α of .90.

The Patient Health Questionnaire-9 (PHQ-9) [36] is a measure of depression severity that consists of 9 items scored on a 4-point Likert scale (range 0-3), with the total score ranging from 0 to 27. The internal consistency of this scale was excellent in this sample, with a Cronbach α of .91.

The symptoms of health anxiety were measured with 2 items from the validated Health Anxiety Inventory [37] and 2 items adapted for the COVID-19 pandemic, with one item measuring the specific fear of being infected by the coronavirus and another item measuring the fear of dying from the coronavirus on a 4-point Likert scale (range 0-3). Internal consistency was acceptable for these health anxiety–related questions, with a Cronbach α of .77.

#### *Information-Seeking Behavior*

Participants were asked to estimate the amount of time they had spent obtaining information regarding the COVID-19 pandemic using various media platforms since the beginning of March 2020, which was the onset of pandemic restrictions in Norway. This included (1) recognized newspapers; (2) television channels; (3) social media; (4) forums, blogs, podcasts, and other online outlets (excluding online newspapers); (5) friends, family, and acquaintances; and (6) other sources. Moreover, active avoidance of COVID-19–related information on all media was measured. These variables were measured on an 8-point Likert scale (range 0-7; 0=never, 7=multiple times per hour). Two new variables were created from the existing media variables. Using online interactive media was defined as using social media, forums, and blogs, and was consequently the sum of these variables. Using traditional media was defined as using television and newspapers, and consisted of the sum of these variables.



XSL•FO
RenderX

### Statistical Analyses

The descriptive analyses of the present data were reported using means and standard deviations. Two educational groups were collapsed due to a low N value, and these groups were also used in multiple regressions. For hypothesis 1, 3 separate multiple regression analyses with the symptoms of anxiety, depression, and health anxiety as criterion variables were conducted. The predictor variables were the extents of information obtained about the COVID-19 pandemic from (1) recognized newspapers; (2) television channels; (3) social media; (4) forums, blogs, podcasts, and other online outlets (excluding online newspapers); (5) friends, family, and acquaintances; and (6) other sources, as well as actively staying away from information. Additionally, we controlled for age, gender, education level, and the presence of a psychiatric diagnosis.

For hypothesis 2, 3 separate multiple regression analyses were conducted, with the symptoms of anxiety, depression, and health anxiety as criterion variables. Media variables were collapsed as described above into traditional media and online interactive media. Additionally, we controlled for age, gender, education level, and the presence of a psychiatric diagnosis.

For all multiple regressions, part correlations were calculated. Part correlation, also referred to as semipartial correlation, is a measure low in bias that is easily interpretable. Part correlations are estimates of the strength of a predictive relationship and can be interpreted as an effect size measure using the Cohen criteria of small effect size >0.10, medium effect size >0.30, and large effect size >0.50 [38,39]. Standard criteria for multicollinearity were fulfilled [40] in all multiple regression models, and all assumptions were met.

The preregistered criterion for significance was set at $P<.01$, given the anticipated sample size. All statistical analyses were conducted using R (version 4.0.3; The R Project for Statistical Computing).

### Poststratification Weights

In this study, the sampled gender, age, education, regional affiliation, and ethnic background deviated somewhat from the population parameters. All these deviations (minor deviations as well as larger deviations) were weighted and adjusted to accurately reflect the Norwegian adult population. More weight was assigned to underrepresented units and less weight to overrepresented units. Weights were calculated using the R packages “anesrake” (version 0.8) and “survey” (version 4.0). These packages use an iterative algorithm (ie, raking ratio estimation) to iteratively assign appropriate weights to each subgroup by turn to avoid the distribution matching of one factor skewing the distribution of other factors. The weights outputted as a result of this algorithm were applied to the data set, resulting in a weighted data set with parameters closely matching that of the true population and yielding a highly representative sample. All statistical analyses utilized this weighted representative sample.

## Results

### Sample Characteristics

A total of 4936 individuals were included in this sample, but as 2 of the subgroups (transgendered and intersex individuals) were too small to be factors in our analyses, a sample of 4921 individuals was used for descriptive statistics (Table 1) and for all multiple regression analyses. Table 1 also shows the weighted N values after poststratification with the population parameters as a reference point. Means and standard deviations for the media variables and avoidance can be found in Multimedia Appendix 1, showing that mean time spent gathering information about COVID-19 was the highest for newspapers, television, and social media. A correlation matrix for media variables can be found in Multimedia Appendix 2, showing a negative correlation of avoidance with all media variables as expected.



XSL•FO
RenderX

**Table 1.** Sample characteristics and weighted characteristics (N=4921).

| Variable | Sample, n (%) | Weighted, n (%) | Actual population, % |
|---|---|---|---|
| **Sex** | | | |
| Female | 3911 (79.48) | 2427 (49.32) | 49.77 |
| Male | 1010 (20.52) | 2494 (50.68) | 50.23 |
| **Age group (years)** | | | |
| 18-30 | 1703 (34.61) | 1069 (21.73) | 23.22 |
| 31-44 | 1606 (32.64) | 1340 (27.24) | 24.30 |
| 45-64 | 1344 (27.31) | 1707 (34.70) | 31.26 |
| 65 or above | 268 (5.44) | 804 (16.34) | 21.22 |
| **Education** | | | |
| Junior high school[a] | 191 (3.88) | 573 (11.64) | 25.40 |
| Completed high school | 736 (14.96) | 1931 (39.25) | 37.00 |
| Currently studying | 775 (15.75) | 409 (8.32) | 6.70 |
| Completed university degree | 3219 (65.41) | 2006 (40.78) | 30.90 |
| **Ethnic background** | | | |
| Native Norwegian | 4563 (92.73) | 4346 (88.32) | 85.29 |
| European | 274 (5.57) | 401 (8.16) | 7.58 |
| Asian | 39 (0.79) | 117 (2.38) | 4.56 |
| African | 6 (0.12) | 18 (0.37) | 1.85 |
| North America/Oceania | 15 (0.30) | 17 (0.35) | 0.27 |
| South/Middle/Latin-America | 24 (0.49) | 21 (0.43) | 0.45 |
| **Regional affiliation** | | | |
| East Norway | 3103 (63.06) | 2966 (60.28) | 58.32 |
| West Norway | 1162 (23.61) | 922 (18.74) | 20.28 |
| Mid-Norway | 482 (9.79) | 781 (15.87) | 15.95 |
| Northern Norway | 174 (3.54) | 251 (5.11) | 5.45 |

[a]This category is collapsed and consists of individuals who did and those who did not complete junior high school.

## Information-Seeking Behavior and the Symptoms of Anxiety

The multiple regression model examining the factors associated with the symptoms of anxiety can be found in Table 2 and Multimedia Appendix 3. Multimedia Appendix 3 displays the regression results for the variables online interactive media and traditional media, explaining 33% of the variance in the data. Gender was a significant predictor of anxiety, with female gender being associated with higher levels of anxiety symptoms. Age was also a significant predictor of anxiety symptoms, with lower age being associated with higher symptoms of anxiety. Additionally, having a pre-existing mental health condition was associated with increased anxiety symptoms. Time spent obtaining information about the pandemic using newspapers was a significant predictor of anxiety symptoms, and more time spent reading newspapers was associated with a higher degree of anxiety symptoms.



XSL•FO
RenderX

**Table 2.** Predictors of anxiety symptoms in the weighted representative sample (N=4921; adjusted $R^2$=0.33).

| Variable | Beta | SE of beta | *P* value | Part correlation (r) |
|---|---|---|---|---|
| Intercept | 5.16 | 0.39 | <.001 | 1.00 |
| Gender[a] | −0.59 | 0.18 | .001 | −0.06 |
| Age | −0.06 | 0.01 | <.001 | −0.18 |
| Education | −0.19 | 0.09 | .03 | −0.04 |
| Mental health condition | 4.34 | 0.29 | <.001 | 0.37 |
| Newspapers | 0.28 | 0.07 | <.001 | 0.08 |
| Television | 0.01 | 0.07 | .86 | 0.00 |
| Social media | 0.24 | 0.07 | <.001 | 0.08 |
| Forums and blogs | 0.14 | 0.09 | .10 | 0.04 |
| Friends and family | −0.06 | 0.10 | .58 | −0.01 |
| Other | 0.04 | 0.08 | .57 | 0.01 |
| Avoidance | 0.36 | 0.08 | <.001 | 0.10 |

[a]Female=0; male=1.

Time spent obtaining information from social media was a significant predictor of anxiety symptoms, with more time spent being associated with a higher degree of anxiety symptoms. Obtaining information using both traditional media and online interactive media was a significant predictor of anxiety, and the relationship was positive. Finally, avoiding information about COVID-19 entirely was a significant predictor of anxiety symptoms, with a higher degree of avoidance being associated with an increased level of anxiety symptoms.

Among these significant associations, part correlations revealed the factors most strongly associated with the symptoms of anxiety while accounting for all other variables. The most prominent factors associated with the symptoms of anxiety included having a pre-existing mental health condition (part correlation=0.37), age (part correlation=−0.18), obtaining information from online interactive media (part correlation=0.10), and avoiding information about the pandemic (part correlation=0.10), all revealing small to medium effect sizes. Lower effect sizes emerged for obtaining information concerning the COVID-19 pandemic from social media (part correlation=0.08), for obtaining information from newspapers (part correlation=0.08), and generally for traditional media (part correlation=0.08) and the effect of gender (part correlation=−0.06).

Obtaining information about the pandemic from television, forums and blogs, family and friends, and other sources was unrelated to the symptoms of anxiety. Education level was also unrelated to the symptoms of anxiety at our significance level.

## Information-Seeking Behavior and the Symptoms of Depression

The multiple regression model examining the factors associated with depressive symptoms is depicted in Table 3 and Multimedia Appendix 4. Multimedia Appendix 4 displays the regression results for the variables online interactive media and traditional media, explaining 36% of the variance in the data. Age, gender, and education level were significant predictors of depressive symptoms, where lower age, lower education level, and a trend toward female gender were associated with a higher level of depressive symptoms. Having a pre-existing mental health condition was associated with a greater level of depressive symptoms. Time spent obtaining information about the pandemic from newspapers and social media was a significant predictor of depressive symptoms, with more time spent reading newspapers being associated with a higher level of depressive symptoms.



XSL•FO
RenderX

**Table 3.** Predictors of depressive symptoms in the weighted representative sample (N=4921; adjusted $R^2$=0.33).

| Variable | Beta | SE of beta | *P* value | Part correlation (r) |
|---|---|---|---|---|
| Intercept | 8.42 | 0.54 | <.001 | 1.00 |
| Gender[a] | −0.68 | 0.25 | .005 | −0.05 |
| Age | −0.07 | 0.01 | <.001 | −0.16 |
| Education | −0.34 | 0.12 | .004 | −0.05 |
| Mental health condition | 6.41 | 0.39 | <.001 | 0.41 |
| Newspapers | 0.27 | 0.09 | .002 | 0.06 |
| Television | −0.12 | 0.09 | .19 | −0.03 |
| Social media | 0.22 | 0.08 | .009 | 0.05 |
| Forums and blogs | 0.18 | 0.11 | .11 | 0.03 |
| Friends and family | −0.25 | 0.13 | .06 | −0.04 |
| Other | −0.02 | 0.10 | .82 | 0.001 |
| Avoidance | 0.64 | 0.11 | <.001 | 0.13 |

[a]Female=0; male=1.

Obtaining information about the pandemic through online interactive media was a significant predictor of depressive symptoms, where more time spent was associated with a higher level of depressive symptoms. Avoiding information entirely was a significant predictor of depressive symptoms, where increasing avoidance of COVID-19–related information was associated with a higher level of depressive symptoms.

Among these significant associations, part correlations revealed the factors most strongly associated with depressive symptoms while accounting for all other variables. The strongest factors associated with depressive symptoms included having a pre-existing mental health condition (part correlation=0.41), age (part correlation=−0.16), and avoiding information about COVID-19 (part correlation=0.13), all revealing small to medium effect sizes. Smaller effect sizes were observed for education level (part correlation=−0.06), and obtaining news about the pandemic through newspapers (part correlation=0.06), online interactive media (part correlation=0.08), and traditional media (part correlation=0.03).

Obtaining information about COVID-19 from television, forums and blogs, family and friends, and other sources was unrelated to depressive symptoms. Obtaining information from traditional media was also unrelated to depressive symptoms.

## Information-Seeking Behavior and the Symptoms of Health Anxiety

The multiple regression model examining the factors associated with the symptoms of health anxiety can be found in Table 4 and Multimedia Appendix 5. Multimedia Appendix 5 displays the regression results for the variables online interactive media and traditional media, and the regressions in Table 4 and Multimedia Appendix 5 explain 16% and 15% of the variance in the data, respectively. Education was a significant predictor of health anxiety symptoms, with a lower education level being associated with a higher level of health anxiety symptoms. Having a pre-existing mental health diagnosis was a significant predictor of health anxiety symptoms.



XSL•FO
RenderX

**Table 4.** Predictors of health anxiety symptoms in the weighted representative sample (N=4921; adjusted $R^2$=0.16).

| Variable | Beta | SE of beta | *P* value | Part correlation (r) |
|---|---|---|---|---|
| Intercept | 0.52 | 0.18 | .003 | 1.00 |
| Gender[a] | −0.13 | 0.09 | .13 | −0.03 |
| Age | 0.00 | 0.00 | .13 | 0.03 |
| Education | −0.19 | 0.04 | <.001 | −0.09 |
| Mental health condition | 1.19 | 0.14 | <.001 | 0.24 |
| Newspapers | 0.20 | 0.03 | <.001 | 0.13 |
| Television | 0.02 | 0.03 | .61 | 0.01 |
| Social media | 0.12 | 0.03 | <.001 | 0.09 |
| Forums and blogs | 0.03 | 0.05 | .49 | 0.02 |
| Friends and family | −0.03 | 0.05 | .61 | −0.01 |
| Other | 0.02 | 0.04 | .69 | 0.01 |
| Avoidance | 0.09 | 0.04 | .007 | 0.06 |

[a]Female=0; male=1.

Time spent obtaining information from newspapers was a significant predictor of health anxiety symptoms, with more time spent being associated with an increased level of health anxiety symptoms. Time spent obtaining information from social media was a significant predictor of health anxiety symptoms, with more time spent being associated with a higher level of health anxiety symptoms. Obtaining information about the pandemic from both traditional media and online interactive media was a significant predictor of health anxiety, where more time spent on these media was associated with increased symptoms. Avoiding information entirely was a significant predictor of health anxiety symptoms, with higher levels of avoidance being positively associated with health anxiety symptoms.

Among these significant associations, part correlations revealed the factors most strongly associated with the symptoms of health anxiety while accounting for all other variables. The strongest factors associated with the symptoms of health anxiety included having a pre-existing mental health condition (part correlation=0.23), using traditional media to obtain information about the pandemic (part correlation=0.14), and time spent gathering information from newspapers (part correlation=0.13), in addition to using online interactive media in general (part correlation=0.10), revealing small effect sizes. Obtaining information through social media (part correlation=0.09), education level (part correlation=−0.09), and avoiding news about COVID-19 (part correlation=0.06) had smaller effect sizes.

Age; gender; and information attainment related to COVID-19 from television, family and friends, forums and blogs, and other sources were unrelated to the symptoms of health anxiety.

## Discussion

### Hypothesized Effects

This study investigated the association of various sources of pandemic information dissemination with psychopathology during the COVID-19 pandemic in a Norwegian sample. For hypothesis 1, the strongest effect sizes were found for using social media and newspapers to obtain information about COVID-19. Support was not found for the hypothesized associations of obtaining information from television, forums and blogs, friends and family, and other sources with the symptoms of anxiety, health anxiety, and depression.

For hypothesis 2, using either online interactive media or traditional media had the strongest effect sizes pertaining to the symptoms of psychopathology, except for the measure of depressive symptoms, where using traditional media was not statistically significant. Additionally, as hypothesized, avoiding news about COVID-19 was significantly associated with the symptoms of anxiety, health anxiety, and depression.

### Information-Seeking Behavior and the Symptoms of Psychopathology

Our findings seem to be in line with previous pandemic research results, where the use of different sources of information about COVID-19 was associated with increased symptoms of anxiety, depression, and health anxiety [12,13,41,42]. During such pandemics, individuals are required to live life with a certain degree of uncertainty due to the novelty of the situation [29], such as living with the risk of getting infected, not knowing if others you interact with are contagious, and being constantly wary of potential situations that could put one at risk.

Individuals with high intolerance to uncertainty may attempt to reduce this uncertainty through behaviors, such as checking the internet (ie, social media, forums, and blogs) and other sources (ie, newspapers and television) for information [29-31], exuding internal control over an external problem. Furthermore, health anxiety is tied to an increase in the search for health-related information, with psychological distress as a potential consequence [32]. With an increased tendency to search for information, there is also an increase in different media providing that information, with potentially differential effects.



XSL•FO
RenderX

Additionally, using social media in general is associated with the symptoms of mental distress [14-19].

As with social media, other unverified platforms have been investigated in previous pandemics. During the 2009 swine flu outbreak, language in blogs mentioning the swine flu was more dramatic compared to a control, and this language later spread to newspapers [43], indicating that monitoring of multiple types of media may result in exposure to similar dramatic language over time across different media. Interestingly, a study by Blakey [44] during the 2015 Zika outbreak found that knowledge of the virus predicted anxiety, indicating that it may not only be dramatic language that has a detrimental effect, but also general knowledge about the outbreak itself, an important point that national and regional governments need to take into consideration when preparing information for their citizens.

### Information Avoidance and the Symptoms of Psychopathology

In line with the findings of this investigation, avoidance in anxiety literature is empirically documented as a maladaptive behavior that maintains anxiety levels [45-50]. Our results indicate that the association of news avoidance with anxiety and depression was greater than the association of the use of different media platforms for information about the pandemic with the symptoms of psychopathology, when viewing part correlations as effect sizes.

One reason for this might be that individuals who explicitly avoid news show general tendencies of avoidance, although existing mental health conditions were controlled for in our analyses. Previous research has shown that those who avoid news intentionally are exposed to news regardless [51-53]. Actively avoiding news and then being exposed to the news regardless might trigger the very feelings one seeks to avoid, although in an uncontrolled context, it might be more hurtful than being exposed to the news in the first place. Avoiding news might be maladaptive in the same sense that avoiding a feared stimuli for a person who has an anxiety disorder hinders recovery. Avoiding such information can shield the individual from information that might trigger existing fearful thoughts, but can possibly increase the span of attentiveness toward general negative experience, which may in turn be reinterpreted as threatening symptom stimuli through attentional biases [54].

Consequently, a core question is how we can reach those avoiding pandemic-related information, as these individuals seem to be at increased risk of experiencing psychopathological symptoms. To better understand the reasons underlying avoidance, future studies should explore whether pandemic-related news avoidance is intentional or unintentional, and how it is related to “news overload.” Skovgaard and Andersen [55] argue that one potential cause for avoidance is the overload of information, which is in accordance with the term infodemic. This connection is further elucidated in recent studies on social media [56,57], showing that news avoidance directly affects news efficacy.

Furthermore, individuals may be engaged in intentional avoidance due to low trust in news and the negativity of news [55]. The authors propose that these problems underlying intentional avoidance may be solved by providing clear and concise information, avoiding presentation of opinions, sticking to reporting facts, and rebalancing negative perspectives with future-oriented solutions to the presented problems. Rebalancing the overload of negative information present during pandemics with solutions on how to tackle the pandemic is also important from a psychological perspective, with the ability to increase the psychological need of competence and self-efficacy [58].

Governments might also use the research on incidental exposure to news to position information on platforms and in contexts where the information is likely to be seen, even in passing, especially considering that there are groups of people who use social media as their primary source of news [41]. As both this study and other studies have shown associations of social media use for news consumption regarding the current pandemic and social media use in general with the symptoms of psychopathology, it becomes vitally important to engage readers to both seek and receive fact-checked information on these media. Another possible solution could be attempted by presenting imperative pandemic-related news along with entertainment programs on television, and in the context of entertainment on the internet, previous studies have shown that people are exposed to this information even in passing [41]. Through these strategies designed to tackle intentional and unintentional avoidance, journalists and health-policy makers may provide avoidant individuals with essential life-saving information, and ease added symptoms of psychopathology associated with avoidance of news related to the pandemic.

### Strengths and Limitations

A major strength of this study is that data collection was conducted in a period where pandemic restrictions had just been lifted, and no significant changes to the pandemic protocols in Norway occurred. Although the collected data were not fully representative of the Norwegian population given the oversampling and undersampling of several highlighted subgroups, poststratification weights were calculated and applied to closely match the sample with the true distribution in the population, yielding a representative sample of the Norwegian adult population. The study used a cross-sectional design, which impairs the ability to infer causal effects. Another limitation is using self-reporting of symptoms and inquiring about the presence of a psychiatric diagnosis rather than performing diagnostic interviews. As with traditional sampling techniques involving voluntary participation, potential self-selection of participants may have occurred. However, self-selection was accounted for statistically through the described weighting and raking procedures.

### Conclusion

This study found significant associations between the symptoms of psychopathology and the use of media for obtaining information related to the COVID-19 pandemic. Significant findings for obtaining information through newspapers, social media, and online interactive media were seen across all 3 measures of psychopathology. Avoidance of COVID-19–related information and associations with the symptoms of psychopathology emerged as core findings, with generally higher effect sizes compared with informational attainment. In



XSL•FO
RenderX

the age of social media and fast-paced informational streams, presenting up-to-date curated pandemic information is imperative to mitigate the associated detrimental mental health consequences of information dissemination, and efficacious communication is important to hinder viral spread, keep populations up to date on the latest hygienic recommendations, and counteract the effects of infodemic misinformation and news overload. Effectively and accurately conveying information can directly benefit the mental health of citizens.

Further research should be conducted to elucidate the causal mechanisms underlying the associations between media use and the symptoms of psychopathology, and between news avoidance and the symptoms of psychopathology. Public health officials must direct their efforts toward creating better opportunities for individuals to acquire validated information related to the pandemic, nudging them to trusted sources of information, and preventing news overload.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Data for media variables.
[DOCX File , 17 KB - formative_v5i12e28239_app1.docx ]

Multimedia Appendix 2
Correlation matrix for media variables.
[DOCX File , 17 KB - formative_v5i12e28239_app2.docx ]

Multimedia Appendix 3
Predictors of anxiety symptoms in the weighted representative sample.
[DOCX File , 17 KB - formative_v5i12e28239_app3.docx ]

Multimedia Appendix 4
Predictors of depressive symptoms in the weighted representative sample.
[DOCX File , 17 KB - formative_v5i12e28239_app4.docx ]

Multimedia Appendix 5
Predictors of health anxiety symptoms in the weighted representative sample.
[DOCX File , 17 KB - formative_v5i12e28239_app5.docx ]

## References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
2. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
3. Covid-19 disrupting mental health services in most countries. World Health Organization. 2020. URL: https://www.who.int/news/item/05-10-2020-covid-19-disrupting-mental-health-services-in-most-countries-who-survey [accessed 2021-11-26]
4. Panda PK, Gupta J, Chowdhury SR, Kumar R, Meena AK, Madaan P, et al. Psychological and Behavioral Impact of Lockdown and Quarantine Measures for COVID-19 Pandemic on Children, Adolescents and Caregivers: A Systematic Review and Meta-Analysis. J Trop Pediatr 2021 Jan 29;67(1):A [FREE Full text] [doi: 10.1093/tropej/fmaa122] [Medline: 33367907]
5. Sahebi A, Nejati-Zarnaqi B, Moayedi S, Yousefi K, Torres M, Golitaleb M. The prevalence of anxiety and depression among healthcare workers during the COVID-19 pandemic: An umbrella review of meta-analyses. Prog Neuropsychopharmacol Biol Psychiatry 2021 Apr 20;107:110247 [FREE Full text] [doi: 10.1016/j.pnpbp.2021.110247] [Medline: 33476692]
6. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. The Lancet Psychiatry 2020 Jun 15;7(6):547-560. [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
7. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
8. Kelly BD. Coronavirus disease: challenges for psychiatry. Br J Psychiatry 2020 Jul 15;217(1):352-353 [FREE Full text] [doi: 10.1192/bjp.2020.86] [Medline: 32293555]



XSL•FO
RenderX

9. Ebrahimi OV, Hoffart A, Johnson SU. Physical Distancing and Mental Health During the COVID-19 Pandemic: Factors Associated With Psychological Symptoms and Adherence to Pandemic Mitigation Strategies. Clinical Psychological Science 2021 Mar 12;9(3):489-506. [doi: 10.1177/2167702621994545]
10. Lang A. Using the Limited Capacity Model of Motivated Mediated Message Processing to Design Effective Cancer Communication Messages. Journal of Communication 2006 Aug;56:S57-S80. [doi: 10.1111/j.1460-2466.2006.00283.x]
11. Heath C. Do People Prefer to Pass Along Good or Bad News? Valence and Relevance of News as Predictors of Transmission Propensity. Organ Behav Hum Decis Process 1996 Nov;68(2):79-94. [doi: 10.1006/obhd.1996.0091] [Medline: 8954872]
12. Cheng C, Ebrahimi OV, Lau Y. Maladaptive coping with the infodemic and sleep disturbance in the COVID-19 pandemic. J Sleep Res 2021 Aug 27;30(4):e13235 [FREE Full text] [doi: 10.1111/jsr.13235] [Medline: 33247519]
13. Taha S, Matheson K, Cronin T, Anisman H. Intolerance of uncertainty, appraisals, coping, and anxiety: the case of the 2009 H1N1 pandemic. Br J Health Psychol 2014 Sep;19(3):592-605. [doi: 10.1111/bjhp.12058] [Medline: 23834735]
14. Ahmad AR, Murad HR. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
15. Chao M, Xue D, Liu T, Yang H, Hall BJ. Media use and acute psychological outcomes during COVID-19 outbreak in China. J Anxiety Disord 2020 Aug;74:102248 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102248] [Medline: 32505918]
16. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. International Journal of Adolescence and Youth 2019 Mar 21;25(1):79-93. [doi: 10.1080/02673843.2019.1590851]
17. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association between social media use and depression among U.S. young adults. Depress Anxiety 2016 Apr;33(4):323-331 [FREE Full text] [doi: 10.1002/da.22466] [Medline: 26783723]
18. Rubin GJ, Amlôt R, Wessely S, Greenberg N. Anxiety, distress and anger among British nationals in Japan following the Fukushima nuclear accident. Br J Psychiatry 2012 Nov;201(5):400-407 [FREE Full text] [doi: 10.1192/bjp.bp.112.111575] [Medline: 22995630]
19. Aalbers G, McNally RJ, Heeren A, de Wit S, Fried EI. Social media and depression symptoms: A network perspective. J Exp Psychol Gen 2019 Aug;148(8):1454-1462. [doi: 10.1037/xge0000528] [Medline: 30507215]
20. Primack BA, Shensa A, Escobar-Viera CG, Barrett EL, Sidani JE, Colditz JB, et al. Use of multiple social media platforms and symptoms of depression and anxiety: A nationally-representative study among U.S. young adults. Computers in Human Behavior 2017 Apr;69:1-9. [doi: 10.1016/j.chb.2016.11.013]
21. Vannucci A, Flannery KM, Ohannessian CM. Social media use and anxiety in emerging adults. J Affect Disord 2017 Jan 01;207:163-166. [doi: 10.1016/j.jad.2016.08.040] [Medline: 27723539]
22. Ivie EJ, Pettitt A, Moses LJ, Allen NB. A meta-analysis of the association between adolescent social media use and depressive symptoms. J Affect Disord 2020 Oct 01;275:165-174. [doi: 10.1016/j.jad.2020.06.014] [Medline: 32734903]
23. Strekalova YA. Health Risk Information Engagement and Amplification on Social Media. Health Educ Behav 2017 Apr;44(2):332-339. [doi: 10.1177/1090198116660310] [Medline: 27413028]
24. Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. World Health Organization. 2020 Sep. URL: https://tinyurl.com/2p996hb4 [accessed 2021-02-25]
25. Chan A, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. Anaesthesia 2020 Dec;75(12):1579-1582 [FREE Full text] [doi: 10.1111/anae.15057] [Medline: 32227594]
26. Pribble JM, Goldstein KM, Fowler EF, Greenberg MJ, Noel SK, Howell JD. Medical news for the public to use? What's on local TV news. Am J Manag Care 2006 Mar;12(3):170-176 [FREE Full text] [Medline: 16524349]
27. Scherr S. Traditional media use and depression in the general population: evidence for a non-linear relationship. Curr Psychol 2018 Oct 05;40(2):957-972. [doi: 10.1007/s12144-018-0020-7]
28. Ebrahimi OV, Hoffart A, Johnson SU. Viral mitigation and the COVID-19 pandemic: factors associated with adherence to social distancing protocols and hygienic behaviour. Psychol Health 2021 Aug 02:1-24. [doi: 10.1080/08870446.2021.1960987] [Medline: 34339328]
29. Taylor S. The Psychology of Pandemics: Preparing for the Next Global Outbreak of Infectious Disease. Newcastle upon Tyne, United Kingdom: Cambridge Scholars Publishing; 2019.
30. Robichaud M, Koerner N, Dugas MJ. Cognitive Behavioral Treatment for Generalized Anxiety Disorder From Science to Practice. New York, NY: Routledge; 2019.
31. Shihata S, McEvoy PM, Mullan BA, Carleton RN. Intolerance of uncertainty in emotional disorders: What uncertainties remain? J Anxiety Disord 2016 Jun;41:115-124. [doi: 10.1016/j.janxdis.2016.05.001] [Medline: 27212227]
32. Baumgartner SE, Hartmann T. The role of health anxiety in online health information search. Cyberpsychol Behav Soc Netw 2011 Oct;14(10):613-618. [doi: 10.1089/cyber.2010.0425] [Medline: 21548797]
33. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational




XSL•FO
RenderX

studies. Ann Intern Med 2007 Oct 16;147(8):573-577 [FREE Full text] [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
34. Stevens GA, Alkema L, Black RE, Boerma JT, Collins GS, Ezzati M, GATHER Working Group. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. PLoS Med 2016 Jun;13(6):e1002056 [FREE Full text] [doi: 10.1371/journal.pmed.1002056] [Medline: 27351744]
35. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
36. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
37. Salkovskis PM, Rimes KA, Warwick HMC, Clark DM. The Health Anxiety Inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. Psychol Med 2002 Jul 15;32(5):843-853. [doi: 10.1017/s0033291702005822] [Medline: 12171378]
38. Dudgeon P. A Comparative Investigation of Confidence Intervals for IndependentVariables in Linear Regression. Multivariate Behav Res 2016 Mar 25;51(2-3):139-153. [doi: 10.1080/00273171.2015.1121372] [Medline: 27015449]
39. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
40. Hocking RR. Methods and applications of linear models. Hoboken, NJ: John Wiley and Sons; 2013.
41. Bendau A, Petzold MB, Pyrkosch L, Mascarell Maricic L, Betzler F, Rogoll J, et al. Correction to: Associations between COVID-19 related media consumption and symptoms of anxiety, depression and COVID-19 related fear in the general population in Germany. Eur Arch Psychiatry Clin Neurosci 2021 Oct 17;271(7):1409-1409 [FREE Full text] [doi: 10.1007/s00406-021-01290-8] [Medline: 34272977]
42. Pan Y, Xin M, Zhang C, Dong W, Fang Y, Wu W, et al. Associations of Mental Health and Personal Preventive Measure Compliance With Exposure to COVID-19 Information During Work Resumption Following the COVID-19 Outbreak in China: Cross-Sectional Survey Study. J Med Internet Res 2020 Oct 08;22(10):e22596 [FREE Full text] [doi: 10.2196/22596] [Medline: 32936776]
43. Tausczik Y, Faasse K, Pennebaker JW, Petrie KJ. Public anxiety and information seeking following the H1N1 outbreak: blogs, newspaper articles, and Wikipedia visits. Health Commun 2012 Aug 09;27(2):179-185. [doi: 10.1080/10410236.2011.571759] [Medline: 21827326]
44. Blakey SM, Abramowitz JS. Psychological Predictors of Health Anxiety in Response to the Zika Virus. J Clin Psychol Med Settings 2017 Dec 23;24(3-4):270-278 [FREE Full text] [doi: 10.1007/s10880-017-9514-y] [Medline: 29063232]
45. Hofmann SG. Cognitive factors that maintain social anxiety disorder: a comprehensive model and its treatment implications. Cogn Behav Ther 2007 Dec;36(4):193-209 [FREE Full text] [doi: 10.1080/16506070701421313] [Medline: 18049945]
46. Wells A, Clark DM. A cognitive model of social phobia. In: Heimberg RG, Liebowitz MR, Hope DA, Schneier FR, editors. Social phobia: Diagnosis, assessment, and treatment. New York, NY: Guilford Press; 1995.
47. Berman NC, Wheaton MG, McGrath P, Abramowitz JS. Predicting anxiety: the role of experiential avoidance and anxiety sensitivity. J Anxiety Disord 2010 Jan;24(1):109-113. [doi: 10.1016/j.janxdis.2009.09.005] [Medline: 19819106]
48. Trew JL. Exploring the roles of approach and avoidance in depression: an integrative model. Clin Psychol Rev 2011 Nov;31(7):1156-1168. [doi: 10.1016/j.cpr.2011.07.007] [Medline: 21855826]
49. Ottenbreit ND, Dobson KS. Avoidance and depression: the construction of the Cognitive–Behavioral Avoidance Scale. Behaviour Research and Therapy 2004 Mar;42(3):293-313. [doi: 10.1016/s0005-7967(03)00140-2]
50. Cribb G, Moulds M, Carter S. Rumination and Experiential Avoidance in Depression. Behav. change 2012 Feb 22;23(3):165-176. [doi: 10.1375/bech.23.3.165]
51. Fletcher R, Nielsen RK. Are people incidentally exposed to news on social media? A comparative analysis. New Media & Society 2017 Aug 17;20(7):2450-2468. [doi: 10.1177/1461444817724170]
52. Boczkowski PJ, Mitchelstein E, Matassi M. “News comes across when I’m in a moment of leisure”: Understanding the practices of incidental news consumption on social media. New Media & Society 2018 Jan 02;20(10):3523-3539. [doi: 10.1177/1461444817750396]
53. Gil de Zúñiga H, Weeks B, Ardèvol-Abreu A. Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics. J Comput-Mediat Comm 2017 Apr 11;22(3):105-123. [doi: 10.1111/jcc4.12185]
54. Bar-Haim Y, Lamy D, Pergamin L, Bakermans-Kranenburg MJ, van IJzendoorn MH. Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. Psychol Bull 2007 Jan;133(1):1-24. [doi: 10.1037/0033-2909.133.1.1] [Medline: 17201568]
55. Skovsgaard M, Andersen K. Conceptualizing News Avoidance: Towards a Shared Understanding of Different Causes and Potential Solutions. Journalism Studies 2019 Nov 11;21(4):459-476. [doi: 10.1080/1461670x.2019.1686410]
56. Park CS. Does Too Much News on Social Media Discourage News Seeking? Mediating Role of News Efficacy Between Perceived News Overload and News Avoidance on Social Media. Social Media + Society 2019 Sep 11;5(3):205630511987295. [doi: 10.1177/2056305119872956]
57. Song H, Jung J, Kim Y. Perceived News Overload and Its Cognitive and Attitudinal Consequences for News Usage in South Korea. Journalism & Mass Communication Quarterly 2016 Dec 15;94(4):1172-1190. [doi: 10.1177/1077699016679975]



XSL•FO
RenderX

58. Deci EL, Ryan RM. Self-determination theory. In: Van Lange PAM, Kruglanski AW, Higgins ET, editors. Handbook of Theories of Social Psychology: Volume 1. Thousand Oaks, CA: Sage Publications Ltd; 2012.

## Abbreviations

**GAD-7:** Generalized Anxiety Disorder-7
**PHQ-9:** Public Health Questionnaire-9
**VMP:** viral mitigation protocol

*Edited by G Eysenbach; submitted 25.02.21; peer-reviewed by I Mircheva, A Roy, J Marconato, T Capote, M Subotic-Kerry; comments to author 13.05.21; revised version received 29.06.21; accepted 18.09.21; published 03.12.21.*

*Please cite as:*
*Amundsen OM, Hoffart A, Johnson SU, Ebrahimi OV*
*Pandemic Information Dissemination and Its Associations With the Symptoms of Mental Distress During the COVID-19 Pandemic: Cross-sectional Study*
*JMIR Form Res 2021;5(12):e28239*
*URL: https://formative.jmir.org/2021/12/e28239*
*doi:10.2196/28239*
*PMID:34678750*

©Ole Myklebust Amundsen, Asle Hoffart, Sverre Urnes Johnson, Omid V Ebrahimi. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Re-examining COVID-19 Self-Reported Symptom Tracking Programs in the United States: Updated Framework Synthesis

Miranda Lynn Janvrin[1,2], MPH; Jessica Korona-Bailey[1,2], MPH; Tracey Pérez Koehlmoos[2], MHA, PhD

[1]The Henry M Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, United States

[2]Uniformed Services University of the Health Sciences, Bethesda, MD, United States

**Corresponding Author:**
Miranda Lynn Janvrin, MPH
The Henry M Jackson Foundation for the Advancement of Military Medicine
6720B Rockledge Dr
Suite 100
Bethesda, MD, 20817
United States
Phone: 1 6035403059
Email: miranda.janvrin@gmail.com

## Abstract

**Background:** Early in the pandemic, in 2020, Koehlmoos et al completed a framework synthesis of currently available self-reported symptom tracking programs for COVID-19. This framework described relevant programs, partners and affiliates, funding, responses, platform, and intended audience, among other considerations.

**Objective:** This study seeks to update the existing framework with the aim of identifying developments in the landscape and highlighting how programs have adapted to changes in pandemic response.

**Methods:** Our team developed a framework to collate information on current COVID-19 self-reported symptom tracking programs using the "best-fit" framework synthesis approach. All programs from the previous study were included to document changes. New programs were discovered using a Google search for target keywords. The time frame for the search for programs ranged from March 1, 2021, to May 6, 2021.

**Results:** We screened 33 programs, of which 8 were included in our final framework synthesis. We identified multiple common data elements, including demographic information such as race, age, gender, and affiliation (all were associated with universities, medical schools, or schools of public health). Dissimilarities included questions regarding vaccination status, vaccine hesitancy, adherence to social distancing, COVID-19 testing, and mental health.

**Conclusions:** At this time, the future of self-reported symptom tracking for COVID-19 is unclear. Some sources have speculated that COVID-19 may become a yearly occurrence much like the flu, and if so, the data that these programs generate is still valuable. However, it is unclear whether the public will maintain the same level of interest in reporting their symptoms on a regular basis if the prevalence of COVID-19 becomes more common.

***(JMIR Form Res 2021;5(12):e31271)*** doi:10.2196/31271

**KEYWORDS**

COVID-19; coronavirus; framework analysis; information resources; monitoring; patient-reported outcome measures; self-reported; surveillance; symptom tracking; synthesis; digital health

## Introduction

Over the past few years, the COVID-19 pandemic has resulted in more than 175 million deaths around the globe, and it has fundamentally changed the lives of millions more, demanding continual innovation and ingenuity. The increasing isolation brought on by the recommended quarantine and social distancing guidelines has led to further reliance on technology for both employment and social interaction [1,2]. Capitalizing on this, researchers in the field of participatory epidemiology developed programs to track individuals' symptoms and map trends within communities.

These developments were further spurred by the delay in traditional epidemiological surveillance. Although epidemiological surveillance is fundamental in coordinating strategies for detection and prevention [3], in the early stages



XSL•FO
RenderX

of the COVID-19 pandemic, there were many instances where testing was not available or otherwise inaccessible [4]. It was postulated that this delay could be partially assuaged through an uptake in crowdsourced participatory surveillance efforts, based on the successful monitoring of yearly influenza outbreaks [5,6]. Although outbreaks have considerably declined since the advent of the COVID-19 vaccine, the application of COVID-19 symptom tracking technology to the pandemic has afforded researchers ample data both on virus spread and its impact on the individual.

To shed light on these important efforts in the United States, early in the pandemic, in 2020, Koehlmoos et al [7] completed a framework synthesis of currently available self-reported symptom tracking programs for COVID-19—the disease caused by a specific novel coronavirus. They sought to identify programs in the United States that remotely tracked daily symptom fluctuations among participants and mapped disease spread at a community level. This framework described the programs, partners and affiliates, funding, responses, platform, and intended audience, among other considerations. Ideally, this synthesis would raise awareness of programs while identifying gaps and overlaps [7]. This study seeks to update Koehlmoos et al's framework, with the aim of identifying developments in the landscape and highlighting how programs have adapted to changes in pandemic response.

# Methods

## Study Design

Our team developed a framework to collate information on current COVID-19 self-reported symptom tracking programs using the "best-fit" framework synthesis approach [8]. The innovative best-fit framework was chosen for its strength, utility, and appropriateness in drawing conclusions for an evolving subject, as it does not require a further process of interpretation by policy makers and practitioners in order for them to inform practice [9]. Information on each program was collected and organized into a table for ease of comparison.

## Target Population

The original framework synthesis served to identify programs tracking symptoms associated with COVID-19 in the US population. The updated version includes all programs identified in the prior synthesis, as well as sought to include any programs implemented thereafter. Inclusion and exclusion criteria are described below.

### *Inclusion Criteria*

Programs that aim to capture and geographically collate self-reported symptoms associated with COVID-19 that are available for use in the United States were included for the analysis. For this synthesis, a symptom tracking program is defined as a program that allows individuals to report COVID-19 symptoms, in an effort to identify geographic areas with emerging disease or changes in disease progression.

### *Exclusion Criteria*

Programs that do not track specific symptoms for COVID-19, those that were symptom checkers intended for individual use only, or those that did not target the US population were excluded from the analysis.

## Program Identification

All programs from the previous synthesis were included to document changes, if any. New programs were discovered using a Google search for keywords (eg, "symptom trackers covid," "symptom trackers coronavirus," "symptom tracking covid," "symptom tracking coronavirus," "daily symptom tracking covid," "daily symptom tracking coronavirus," "self-reporting covid," "self-reporting coronavirus"). The time frame for the search for programs ranged from March 1, 2021, to May 6, 2021. This 2-month timeline was selected to reflect the state of COVID-19 symptom tracking programs 1 year after our previous synthesis.

## Screening Method

Reviewers (JK, MLJ, and TPK) screened programs to determine whether inclusion criteria were met, and if so, they extracted data from program websites using a standardized form. In the case of conflict between reviewers, final determinations were made by TPK. To gather information that may not be available via program websites, program managers were contacted via email.

## Synthesis Method

Data related to program characteristics were extracted from all included programs and organized into a table using the tool developed in the previous synthesis to guide data collection and build the framework for analysis. Data were then synthesized to form meaningful statements about the programs.

# Results

We identified 33 programs in total, 8 (24%) of which met the inclusion criteria. All 6 programs identified in the last synthesis were included to document changes, if any. Notably, 2 of those 6 programs had been rebranded since our original synthesis, with COVIDNearYou now known as Outbreaks Near Me and COVID Symptom Tracker now known as COVID Symptom Study. Two newly identified programs, COVIDSymptom and COVID Control, met the inclusion criteria. Thus, information was gathered from all 8 eligible programs (ie, BeatCOVID19Now, COVIDcast, COVID Control, COVIDSymptom, COVID Symptom Study, HelpBeatCOVID19, HowWeFeel, and Outbreaks Near Me; see Table 1). In all, 25 programs were excluded as they did not meet the inclusion criteria (Multimedia Appendix 1). Of the original 6 programs, BeatCOVID19Now has since been terminated, reportedly due to control of disease spread in its home location (Australia), resulting in lack of interest in continuing efforts in the United States and other international settings. Hence, BeatCOVID19Now is not included in the tabulated data.



XSL•FO
RenderX

**Table 1.** Overview of self-reported symptom tracker programs.

| Characteristic | Program name | | | | | | |
|---|---|---|---|---|---|---|---|
| | COVIDcast | COVID Control | COVIDSymptom | COVID Symptom Study[a] | HelpBeat-COVID19 | HowWeFeel | Outbreaks Near Me[b] |
| Host institution and partners | • Carnegie Mellon University Delphi Research Group<br>• Facebook<br>• COVID Symptom Study[c]<br>• Outbreaks Near Mec | • John Hopkins Bloomberg School of Public Health<br>• John Hopkins School of Medicine<br>• John Hopkins Whiting School of Engineering<br>• John Hopkins Medicine Technology Innovation Center<br>• University of Vermont Medical Center<br>• Capitol Technology University<br>• ITC Infotech Digital Experience | • University of Michigan<br>• Kirusa<br>• NJ Tech Council<br>• NJ Business & Industry Association<br>• Walk-In Medical Urgent Care<br>• Sills, Cummis & Gross<br>• Decagon<br>• Strand<br>• SpectraMEDI | • Harvard T.H. Chan School of Public Health<br>• Massachusetts General Hospital<br>• King's College London<br>• Stanford University School of Medicine<br>• ZOE<br>• COVIDcast[c]<br>• Outbreaks Near Mec<br>• Agricultural Health Study[c]<br>• American Cancer Society[c]<br>• Aspree XT[c]<br>• Black Women's Health Study[c]<br>• California Teachers Study[c]<br>• Cancer Prevention Study–3[c]<br>• Dr. Susan Lovec Foundation<br>• Gulf Study[c]<br>• Gutsc<br>• The Multiethnic Cohort Study[c]<br>• The Sister Study[c]<br>• Stand Up to Cancer[c]<br>• University of Texas School of Public Health[c] | • University of Alabama<br>• Alabama Department of Public Health | • Harvard T.H Chan School of Public Health<br>• MIT[d]<br>• IQSS[e]<br>• McGovern Institute<br>• Howard Hughes Medical Institute<br>• Weizmann Institute of Science<br>• Pinterest<br>• Feeding America<br>• Alex's Lemonade Stand<br>• Chartio<br>• Bill & Melinda Gates Foundation<br>• Stanford University[c]<br>• University of Pennsylvania[c] | • Harvard Medical School<br>• Boston Children's Hospital<br>• Ending Pandemics<br>• Google<br>• Centers for Disease Control<br>• COVID Symptom Study[c]<br>• COVIDcast[c] |
| Location | • Pittsburgh, PA | • Baltimore, MD | • Ann Arbor, MI | • Boston, MA | • Birmingham, AL | • Boston, MA | • Boston, MA |



XSL•FO
RenderX

| Characteristic | Program name | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | COVIDcast | COVID Control | COVIDSymptom | COVID Symptom Study[a] | HelpBeat-COVID19 | HowWeFeel | Outbreaks Near Me[b] |
| Funding sources | • None | • White and Case<br>• ITC Infotech | • New Jersey Technology Control<br>• Kirusa | • Mass General<br>• Wellcome Trust (UK) | • University of Alabama | • Bill and Melinda Gates Foundation<br>• Crowdsourcing | • Ending Pandemics<br>• Crowdsourcing<br>• AtScale[c]<br>• AWS[c,f]<br>• BlazeMeter[c]<br>• Cloudflare[c]<br>• Datadog[c]<br>• MongoDB[c]<br>• SurveyMonkey[c]<br>• TechSoup[c] |
| Intended participants (age in years) | • US residents (18+) | • US residents (13+) | • Worldwide (18+) | • US residents (18+)<br>• Participants from other internal studies, including RCTs[g] | • US residents (18+); particular focus on Alabama and neighboring states | • US residents (18+) | • United States, Canada, and Mexico residents (18+) |
| Date symptom tracker was initiated | • April 2020 | • April 2020 | • June 2020 | • April 2020 | • Not available | • April 2020 | • March 2020 |
| Number of responses to date[c] | • 19,989,000 (+17,415,000)[c] | • 215,000 | —[h] | • 4,651,000 (+4,553,000)[c] | • 101,000 (+44,000)[c] | — | • 5,867,000 (+5,813,000)[c] |
| Mechanism of recruiting participants or platform | • Survey via Facebook | • Apple App Store<br>• Google Play Store | • Apple App Store<br>• Google Play Store | • Apple App Store<br>• Google Play Store | • Web browser | • Apple App Store<br>• Google Play Store | • Web browser |
| Follow-up | • Daily survey prompts via Facebook[c] | • Daily phone notifications | • Daily phone notifications | • Daily phone notifications | • SMS notifications every three days | • Customizable phone notifications[c] | • Daily SMS notifications[c] |
| Frequency of reporting | • Daily | • Daily | — | • Live Data | • Live Data | • Daily | • Weekly |
| Availability of summary tables for external synthesis or utilization | • Yes | • Yes | • Yes | • Yes | • No | • No | • Yes |
| Intended audience for the product | | • Public at large<br>• State and local health officials<br>• Researchers | • Public at large<br>• Health care providers<br>• Researchers | | • Public at large<br>• Neighboring States<br>• State and local health officials<br>• Local policy makers | | |



XSL•FO
RenderX

| Characteristic | Program name | | | | | | |
|---|---|---|---|---|---|---|---|
| | COVIDcast | COVID Control | COVIDSymptom | COVID Symptom Study[a] | HelpBeat-COVID19 | HowWeFeel | Outbreaks Near Me[b] |
| | • Public at large<br>• State and local public health officials<br>• US policy makers<br>• Health care providers<br>• Health care systems<br>• CDCi and national public health organizations[c]<br>• Researchers[c] | | | • Public at large<br>• Participants of internal studies<br>• CDC and national public health organizations[c]<br>• Health care providers[c]<br>• Health care systems[c]<br>• Researchers[c]<br>• State and local public health officials[c]<br>• US policy makers[c] | | • Public at large<br>• State and local public health officials<br>• Health care providers<br>• Health care systems<br>• Researchers<br>• State and local public health professionals<br>• CDC and national public health organizations | • Public at large<br>• CDC and national public health organizations<br>• State and local public health officials<br>• Researchers<br>• Health care providers<br>• Health care systems<br>• US policy makers |
| Publicly available data privacy statement | • Yes | • Yes | • Yes | • Yes | • Yes | • Yes | • Yes |

[a]Formerly known as COVID Symptom Tracker.

[b]Formerly known as COVIDNearYou.

[c]Indicates changes made since the previous synthesis.

[d]MIT: Massachusetts Institute of Technology.

[e]IQSS: Institute for Quantitative Social Science.

[f]AWS: Amazon Web Services.

[g]RCT: randomized controlled trial.

[h]Not available.

[i]CDC: Centers of Disease Control and Prevention.

All 6 of the previously reported programs have maintained their affiliation with a university, school of medicine, or school of public health. The 2 newly added programs are also affiliated with universities, further emphasizing the importance of academic institutions in studying the spread of COVID-19 in the United States. Three programs continue to be based in Boston, MA (ie, COVID Symptom Tracker, HowWeFeel, and Outbreaks Near Me), with another affiliated through partnerships (COVIDcast). However, the 2 new programs are based in Ann Arbor, MI and Baltimore, MD.

All of the programs continued to receive responses through electronic reporting mechanisms, with 2 programs utilizing a web browser-based approach (ie, Outbreaks Near Me and HelpBeatCOVID19). Four programs, including both the newly added programs, use apps for both Apple and Android devices; these include COVID Control, COVIDSymptom, COVID Symptom Study, and HowWeFeel. One program (COVIDcast) uses social media as a platform for their survey.

Upon reviewing the data elements being collected across the programs (Table 2), it appears that some of the trackers have been revised to ask more timely questions, including those covering topics like vaccination (eg, COVIDcast, COVID Symptom Study, HowWeFeel, Outbreaks Near Me), vaccine hesitancy (eg, COVIDcast and HowWeFeel), adherence to social distancing recommendations put forth by the Centers for Disease Control and Prevention (CDC; eg, COVIDcast and HowWeFeel), and in-person schooling for children K-12 (eg, HowWeFeel). Notably, COVIDcast asks about mask usage. The remaining programs (COVID Control, COVIDSymptom, HelpBeatCOVID19, and Outbreaks Near Me) continue to focus their efforts only on elements such as ongoing symptoms, testing status, and demographic information. Interestingly, COVIDcast and COVID Symptom Tracker have limited or removed much of the demographic data that they collect.



XSL•FO
RenderX

**Table 2.** Data elements across programs.

| Question | COVIDcast | COVID Control | COVID Symptom | COVID Symptom Study | HelpBeat-COVID19 | HowWeFeel | Outbreaks Near Me |
|---|---|---|---|---|---|---|---|
| Are you completing the survey on behalf of someone else? | | | | ✓ | ✓ | | |
| Age | | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Gender | | ✓ | ✓ | | ✓ | ✓ | ✓ |
| Zip code | | ✓ | | | ✓ | ✓ | ✓ |
| Race or ethnicity | | | | ✓ | ✓ | ✓ | |
| Blood group | | | | ✓ | | | |
| Symptoms at time of check-in | | ✓ | ✓ | ✓ | | ✓ | |
| Symptoms within the last 24 hours | ✓ | | | | | | |
| Symptoms over the past 7 days | | | | | ✓ | | ✓ |
| Date of symptom onset | | | | | | | ✓ |
| Do others in your household have similar symptoms to those you reported? | ✓ | | | | ✓ | ✓ | |
| Temperature at the time of check-in | | ✓ | | | | | |
| Highest temperature over symptom duration | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Hours of sleep previous night | | | | | | ✓ | |
| Have you been tested for COVID-19? | | ✓ | ✓ | ✓ | ✓ | | |
| Type of testing or care sought due to symptoms | | | | | | ✓ | ✓ |
| What type of medical test did you receive | | | | | | | ✓ |
| Results of testing | | ✓ | ✓ | ✓ | ✓ | | ✓ |
| In the past 14 days, did you want a COVID-19 test but did not receive one | | | | | | | ✓ |
| How long after you started feeling ill did you see a health professional? | | | | | | | ✓ |
| What prescription, if any, did you receive for your illness? | | | | | | | ✓ |
| Have you received the COVID-19 vaccine? | ✓ | | | ✓ | | ✓ | ✓ |
| Would you accept a COVID-19 vaccine if offered? | ✓ | | | ✓ | | | |
| If a safe, effective coronavirus vaccine were available, how likely would you be to get vaccinated? | | | | | | ✓ | |
| What is the main reason you got the COVID-19 vaccine? | | | | | | ✓ | |
| What is the main reason you did not get the COVID-19 vaccine? | | | | | | ✓ | |
| Do you know anyone who has received a COVID-19 vaccine? | | | | | | ✓ | |
| If you have a child under 18, how likely are you to get your child vaccinated? | | | | | | ✓ | |
| Are you experiencing any symptoms near the injection site? | | | | ✓ | | | |
| Preexisting conditions | | | ✓ | ✓ | ✓ | ✓ | |
| Obesity | | | | | ✓ | ✓ | |
| Are you pregnant? | | | | | ✓ | ✓ | |



XSL•FO
RenderX

| Question | COVIDcast | COVID Control | COVID Symptom | COVID Symptom Study | HelpBeat-COVID19 | HowWeFeel | Outbreaks Near Me |
|---|---|---|---|---|---|---|---|
| Are you/have you ever been a smoker | | | ✓ | | ✓ | ✓ | |
| Have you received the flu vaccine? | | | | | | | ✓ |
| Number of people in the household | | | | | ✓ | ✓ | |
| Are you a parent? | | | | | | ✓ | |
| Number of children in the household | | | | | | ✓ | |
| Do any children in your household (pre-K through grade 12) go to full-time in-person classes? | | | | | | ✓ | |
| Has anyone in your household been diagnosed with COVID-19? | | | | | ✓ | | |
| Type of domicile | | | | | ✓ | | |
| Have you left your home in the past 24 hours? | | | | | | ✓ | |
| Reason for leaving home | | | | | | ✓ | |
| What protective measures did you take when you left home? | | | | | | ✓ | |
| In the past 7 days, did you wear a mask most or all of the time in public? | ✓ | | | | | | |
| In the past 7 days, when you were in public places where social distancing is not possible, did most or all other people wear masks? | ✓ | | | | | | |
| In the past 24 hours, did you spend time indoors with someone who isn't currently staying with you? | ✓ | | | | | | |
| In the past 24 hours, did you attend an indoor event with more than 10 people? | ✓ | | | | | | |
| In the past 24 hours, did you go to an indoor market, grocery store, or pharmacy? | ✓ | | | | | ✓ | |
| In the past 24 hours, did you have a meal or drink indoors at a bar, restaurant, or cafe? | ✓ | | | | | ✓ | |
| In the past 24 hours, did you use public transit? | ✓ | | | | | | |
| In the past 7 days, have you traveled outside of your state? | ✓ | | | | | | |
| Have you been in contact with anyone diagnosed with COVID-19? | | | | | ✓ | ✓ | |
| Do you personally know someone in your local community who has COVID-like symptoms? | ✓ | | | | | | |
| Does your job require you to leave your home and go to another place to work where you come in contact with public? | ✓ | | | | ✓ | | |
| How much of the following feelings (tired, calm, happy, angry, sad, thoughtful, optimistic, anxious, lonely, grateful, hopeful, stressed) have you felt so far today? | ✓ | | | | | ✓ | |
| Do you feel very or somewhat worried about becoming seriously ill from COVID-19? | ✓ | | | | | | |



XSL•FO
RenderX

| Question | COVIDcast | COVID Control | COVID Symptom | COVID Symptom Study | HelpBeat-COVID19 | HowWeFeel | Outbreaks Near Me |
|---|---|---|---|---|---|---|---|
| Do you feel very or somewhat worried about your household's finances for the next month? | ✓ | | | | | | |
| Health insurance status | | | | | ✓ | | |
| If you needed to get to a hospital or testing center, how would you get there? | | | | | ✓ | | |
| Can you afford any payment or co-payment required for services? | | | | | ✓ | | |

All 7 currently active programs collect information regarding symptoms, although the time period varies, with the majority (4/7, 57%), asking participants to report current symptoms at the time of check-in (ie, COVID Control, COVIDSymptom, COVID Symptom Tracker, and HowWeFeel), 2 (29%) asking about symptoms experienced over the past week (ie, HelpBeatCOVID19 and Outbreaks Near Me), and only 1 (14%) asking about symptoms experienced over the past 24 hours (ie, COVIDcast). Except for COVIDcast, all other programs ask symptomatic participants who report fever for their temperature, with COVID Control uniquely asking participants who report to have experienced no symptoms. All of these programs (excluding COVIDcast) also ask participants whether they have undergone testing, and 5 programs (ie, COVID Control, COVIDSymptom, COVID Symptom Tracker, HelpBeatCOVID19, Outbreaks Near Me) ask about the results.

COVID Control and COVIDSymptom limited their questions to symptoms, testing, and demographic data, whereas other programs had unique questions, indicating their areas of focus. COVIDcast tailored its questions to current issues, including vaccination status, adherence to social distancing guidelines, mask usage, use of public transit, travel, and mental health concerns. COVID Symptom Study also asked participants about vaccination status, as well as vaccine hesitancy, vaccine side effects, and it was the only program to ask participants about their blood group. HelpBeatCOVID19 asked participants about their household, health insurance status, and financial insecurity. HowWeFeel asked questions on vaccination status; vaccine hesitancy, including asking participants why they have or have not been vaccinated and whether they would vaccinate their children; adherence to social distancing guidelines; and mental health concerns. Finally, Outbreaks Near Me asks participants in-depth questions about testing, including what kind of test participants receive, their diagnosis, and any medications they are prescribed.

## Discussion

### Principal Findings

Over the course of the past year, symptom tracking programs have been exceedingly useful for predictive modeling and population research throughout the COVID-19 pandemic [10]. The data from COVID Symptom Study, for example, has already been used to unearth COVID-19–related trends, even revealing 6 distinct "types" of COVID-19 emerging among their participants. The importance of their contributions were highlighted in the *New England Journal of Medicine* article "Putting the Public Back in Public Health—Surveying Symptoms of Covid-19" [10]. This framework update shows the increase in responses, collaboration, and evolution of questionnaires between programs exemplifying their use by the public health community.

Many symptom checking programs (not included in this framework), defined—for the purposes of this study—as programs that serve to indicate whether the participant likely should or should not seek medical attention without aiming to record symptoms over time, have emerged and with good reason [11]. These programs are critical in deterring those without serious symptoms from seeking emergency care. However, the number of symptom tracking programs has not correspondingly increased. Only 2 additional symptom tracking programs were identified, which indicates either the demand for these programs is significantly lower than symptom checking programs or that the current programs are doing well to serve this niche. It is also possible that both reasons are true.

Reported responses to each of the original 6 programs substantially increased. The lowest number of responses, defined as unique symptom entries by an individual, were received for HelpBeatCOVID19—recorded at 101,000, an increase of 44,000. The remaining 3 programs with reported responses also showed exponential increase in the number of responses. For instance, COVIDcast had 19,989,000 responses, an increase of 17,415,000—the highest of all programs. COVID Symptom Study recorded 4,651,000 responses, an increase by 4,553,000. Outbreaks Near Me recorded 5,867,000 responses, an increase of 5,813,000. In the original framework, two-thirds of the programs had less than 100,000 responses. Among the newly added programs, COVID Control had 215,000 responses. Although only half of the previous programs focused on user retention through notifications (ie, COVID Symptom Study, Outbreaks Near Me, and HelpBeatCOVID19), we found that all programs now prompted previous participants on a daily or semiregular basis through either phone or social media notifications or SMS text messages.

Despite the addition of 2 new programs to the framework, 1 program concluded during the past year. BeatCOVID19Now, based in Australia, concluded and terminated its website in early 2021. Notably, it was the only program in the original synthesis that served international participants and was founded by the same researchers who developed the Influenza Intensity and



XSL•FO
RenderX

Impact Questionnaire (fluiiQ) [12]. The discontinuation of BeatCOVID19Now also raises the question of the longevity of the other programs—like BeatCOVID19Now, Outbreaks Near Me was developed based on existing infrastructure, namely Flu Near You, which performed a similar function for yearly influenza outbreaks. Therefore, discontinuation of BeatCOVID19Now poses the question of whether other programs may be discontinued in the near future.

Collaboration between programs also seems to be auspicious for growth and user retention. Three of the programs (ie, COVIDcast, COVID Symptom Tracker, and Outbreaks Near Me) have begun collaborating. Lack of collaboration was a chief concern in the first iteration of our framework synthesis, with the concern being that it would lead to unnecessary duplication and a division of the pool of potential participants, and ultimately, all of the programs sought to share this important data with those who could best utilize it. Although these concerns cannot be completely assuaged without further collaboration among groups, initiating collaboration among the 3 programs with the highest response rates is a step forward.

A notable development was the continued change in data elements observed by the majority of programs. Elements such as symptom status and potential contact continue to be necessary; however, seeing programs adapt to changes in pandemic response by adding elements concerning vaccination, vaccine hesitancy, social distancing, and mental health indicates what could be the next direction for these programs.

## Limitations

As with the first iteration of this study, several limitations must be acknowledged. First, our analysis was limited to English-language programs; therefore, we may have missed nuances of data collection that are more important to non–English-speaking residents. Second, although the speed of framework analysis enables rapid evaluation of commonalities, it does not provide the in-depth rigor of a full systematic review. Third, our collected data evaluated differences in the number of responses to each program, but it only pertained to what data was collected, how that data is being collected, and what groups created and funded each program. Analyzing the effectiveness, market penetration, or user demographics of evaluated programs is beyond the scope of this study. This kind of analysis, if undertaken in a future study, may provide insight into the question of potential longevity posed by this framework. Fourth, we recognize that program participation is limited to only those who have access to the internet or cellular phone service, creating an unintended disparity among respondents based on their access to and utilization of technology. Therefore, the underlying reasons for the difference in response rate remain beyond the scope of this study. Last, this synthesis neither provides critical appraisal of programs nor evaluates programs for effectiveness.

## Conclusions

At this time, the future of self-reported symptom tracking for COVID-19 and for these programs, is unclear. Some sources have speculated that COVID-19 may become a yearly occurrence much like the flu, and if so, the data that these programs generate is still valuable [13]. However, it is unclear whether the public will maintain the same level of interest in reporting their symptoms on a regular basis if the prevalence of COVID-19 becomes more common, and, for programs like COVIDcast, whether the platform where the survey is hosted will continue to support them. The aforementioned conclusion of BeatCOVID19Now represents a foreboding that could be lying ahead—that lack of interest after virus control through the roll-out of rapid and widely available testing and successful vaccination could lead to a lack of interest among participants, and ultimately discontinuation; however, as of May 2021, these 7 programs have not shown any signs of waning.

## Acknowledgments

The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views, assertions, opinions or policies of the Uniformed Services University of the Health Sciences (USUHS); the Henry M Jackson Foundation for the Advancement of Military Medicine, Inc (HJF); the Department of Defense (DoD); or the Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the US Government.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Programs excluded from the study.
[PNG File , 116 KB - formative_v5i12e31271_app1.png ]

## References

1. Shah SGS, Nogueras D, van Woerden HC, Kiparoglou V. The COVID-19 pandemic: a pandemic of lockdown loneliness and the role of digital technology. J Med Internet Res 2020 Nov 05;22(11):e22287 [FREE Full text] [doi: 10.2196/22287] [Medline: 33108313]
2. De' R, Pandey N, Pal A. Impact of digital surge during Covid-19 pandemic: a viewpoint on research and practice. Int J Inf Manage 2020 Dec;55:102171 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102171] [Medline: 32836633]



XSL•FO
RenderX

3. Burrell C, Howard C, Murphy F. Chapter 14 - Control, Prevention,Eradication. In: Burrell CJ, Howard CR, Murphy FA, editors. Fenner and White's Medical Virology (Fifth Edition). London: Academic Press; 2017:205-216.
4. What we know about delays in coronavirus testing. The Washington Post. 2020 Apr 18. URL: https://www.washingtonpost.com/investigations/2020/04/18/timeline-coronavirus-testing/ [accessed 2021-11-30]
5. Baltrusaitis K, Brownstein JS, Scarpino SV, Bakota E, Crawley AW, Conidi G, et al. Comparison of crowd-sourced, electronic health records based, and traditional health-care based influenza-tracking systems at multiple spatial resolutions in the United States of America. BMC Infect Dis 2018 Aug 15;18(1):403 [FREE Full text] [doi: 10.1186/s12879-018-3322-3] [Medline: 30111305]
6. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu Near You: crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health 2015 Oct;105(10):2124-2130. [doi: 10.2105/ajph.2015.302696]
7. Koehlmoos TP, Janvrin ML, Korona-Bailey J, Madsen C, Sturdivant R. COVID-19 self-reported symptom tracking programs in the United States: framework synthesis. J Med Internet Res 2020 Oct 22;22(10):e23297 [FREE Full text] [doi: 10.2196/23297] [Medline: 33006943]
8. Carroll C, Booth A, Cooper K. A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents. BMC Med Res Methodol 2011 Mar 16;11(1):29 [FREE Full text] [doi: 10.1186/1471-2288-11-29] [Medline: 21410933]
9. Brunton G, Oliver S, Thomas J. Innovations in framework synthesis as a systematic review method. Res Synth Methods 2020 May;11(3):316-330. [doi: 10.1002/jrsm.1399] [Medline: 32090479]
10. Chan AT, Brownstein JS. Putting the public back in public health — surveying symptoms of Covid-19. N Engl J Med 2020 Aug 13;383(7):e45. [doi: 10.1056/nejmp2016259]
11. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
12. Osborne RH, Norquist JM, Elsworth GR, Busija L, Mehta V, Herring T, et al. Development and validation of the Influenza Intensity and Impact Questionnaire (FluiiQ™). Value Health 2011 Jul;14(5):687-699 [FREE Full text] [doi: 10.1016/j.jval.2010.12.005] [Medline: 21839407]
13. Phillips N. The coronavirus is here to stay — here's what that means. Nature. Feb 16. URL: https://www.nature.com/articles/d41586-021-00396-2 [accessed 2021-11-30]

## Abbreviations

**DoD:** Department of Defense
**fluiiQ:** Influenza Intensity and Impact Questionnaire
**HJF:** Henry M Jackson Foundation for the Advancement of Military Medicine, Inc
**USUH:** Uniformed Services University of the Health Sciences

*Edited by G Eysenbach; submitted 15.06.21; peer-reviewed by M Popovic, S Bhattacharjee; comments to author 29.08.21; revised version received 19.09.21; accepted 06.10.21; published 06.12.21.*

*Please cite as:*
*Janvrin ML, Korona-Bailey J, Koehlmoos TP*
*Re-examining COVID-19 Self-Reported Symptom Tracking Programs in the United States: Updated Framework Synthesis*
*JMIR Form Res 2021;5(12):e31271*
*URL: https://formative.jmir.org/2021/12/e31271*
*doi:10.2196/31271*
*PMID:34792469*

©Miranda Lynn Janvrin, Jessica Korona-Bailey, Tracey Pérez Koehlmoos. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Impact of the COVID-19 Pandemic on a Physician Group's WhatsApp Chat: Qualitative Content Analysis

Sawsan Abdel-Razig[1,2], MD, MEHP; Pascale Anglade[1,2], MD, MBA; Halah Ibrahim[3,4], MD, MEHP

[1]Cleveland Clinic Abu Dhabi, Abu Dhabi, United Arab Emirates

[2]Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Clevelend, OH, United States

[3]Department of Medicine, Sheikh Khalifa Medical City, Abu Dhabi, United Arab Emirates

[4]Johns Hopkins University, Baltimore, MD, United States

**Corresponding Author:**
Sawsan Abdel-Razig, MD, MEHP
Cleveland Clinic Abu Dhabi
PO box 112412
Abu Dhabi
United Arab Emirates
Phone: 971 25019999 ext 48460
Email: razigs@clevelandclinicabudhabi.ae

## Abstract

**Background:** Social media has emerged as an effective means of information sharing and community building among health professionals. The utility of these platforms is likely heightened during times of health system crises and global uncertainty. Studies have demonstrated that physicians' social media platforms serve to bridge the gap of information between on-the-ground experiences of health care workers and emerging knowledge.

**Objective:** The primary aim of this study was to characterize the use of a physician WhatsApp (WhatsApp LLC) group chat during the early months of the COVID-19 pandemic.

**Methods:** Through the lens of the social network theory, we performed a qualitative content analysis of the posts of a women physician WhatsApp group located in the United Arab Emirates between February 1, 2020, and May 31, 2020, that is, during the initial surge of COVID-19 cases.

**Results:** There were 6101 posts during the study period, which reflected a 2.6-fold increase in platform use when compared with platform use in the year prior. A total of 8 themes and 9 subthemes were described. The top 3 uses of the platform were requests for information (posts: 2818/6101, 46.2%), member support and promotion (posts: 988/6101, 16.2%), and information sharing (posts: 896/6101, 14.7%). A substantial proportion of posts were related to COVID-19 (2653/6101, 43.5%), with the most popular theme being requests for logistical (nonmedical) information. Among posts containing COVID-19–related medical information, it was notable that two-thirds (571/868, 65.8%) of these posts were from public mass media or unverified sources.

**Conclusions:** Health crises can potentiate the use of social media platforms among physicians. This reflects physicians' tendency to turn to these platforms for information sharing and community building purposes. However, important questions remain regarding the accuracy and credibility of the information shared. Our findings suggest that the training of physicians in social media practices and information dissemination may be needed.

***(JMIR Form Res 2021;5(12):e31791)*** doi:10.2196/31791

**KEYWORDS**

WhatsApp; social media; physician; pandemic; COVID-19; qualitative; communication; misinformation; information-seeking behavior; information seeking; information sharing; content analysis; community

## Introduction

### Background

The COVID-19 pandemic has put an unprecedented and prolonged strain on health systems and health care providers globally. Clinicians are inundated with global developments, an incessant news cycle, and minute-by-minute information from various sources, as new and sometimes conflicting data are becoming available worldwide at an unparalleled pace [1]. The evolving and shifting nature of public health policies,



XSL•FO
RenderX

including curfews, lockdowns, social distancing restrictions, and testing and tracing requirements, presents additional challenges. In order to ensure their personal safety and provide care, frontline medical professionals need to be equipped with the most evidence-based clinical pathways and public health protocols.

### Use of Social Media

Several studies have shown that physicians' social media platforms have become significant facilitators of bridging gaps of information between on-the-ground experiences of health care workers and emerging, scientific, clinical, and population-level knowledge [2,3]. Researchers have analyzed various social media platforms to better understand their use in public health discourse [4,5]. Moreover, infoveillance studies have confirmed a marked increase in individuals' activity on social media platforms, particularly during the COVID-19 pandemic [6]. For example, a COVID-19 physician group, which was created on Facebook in March 2020 and was described as "an inclusive resource for physicians to share front line clinical information about COVID-19 as it becomes available," quickly rose to considerable popularity; the group has approximately 29,000 members to date [7]. Other studies have described how health care providers worldwide have used Twitter and WhatsApp (WhatsApp LLC) during the pandemic to disseminate news and discoveries to colleagues and communicate health information directly to patients [2,8]. There is limited published information however on the use of social media among groups of physicians during a medical and public health crisis.

We previously reported an analysis of the WhatsApp posts of a women physician group; we noted that the platform was effective in enabling female physicians to expand networks, exchange ideas, share scientific information, celebrate accomplishments, and provide support to colleagues [9]. In this study, by using social network theory as an overarching lens [10], we sought to analyze the content of the social media interactions of this group's members during a public health crisis. The primary purpose of this study was to characterize the use of a social media platform among members of a physician group during the COVID-19 pandemic. Our additional aims included examining group members' levels of engagement (ie, by comparing them to members' prepandemic levels of engagement) and identifying the sources of medical information that were shared among the physician members.

## Methods

### Setting and Population

On January 23, 2020, the United Arab Emirates (UAE) reported its first confirmed case of COVID-19. Over the following 4 months, the country experienced a surge of cases; over 60,000 UAE patients were infected with SARS-CoV-2.

WONDER (Women Doctors in the Emirates) is a multispecialty women physician group that originated in 2015 to foster support and collaboration among female physicians living and working in the UAE. In early 2018, WhatsApp Messenger (WhatsApp LLC)—a closed-group messaging app—was added as an adjunct to face-to-face meetings. At the time of writing this paper, group membership totaled 161 physicians, and over 80% (130/161, 80.7%) of members wrote posts during the study period.

### Data Collection and Analysis

Posts from February 1, 2020, through May 31, 2020, were included in the analysis. February 1, 2020, corresponds to the national index case of COVID-19, and the studied period encompasses the initial spring 2020 COVID-19 surge in the UAE, during which the number of COVID-19 cases and related hospitalizations peaked [11]. Prior to data extraction and analysis, group members were informed about this retrospective study via a WhatsApp message and were given the opportunity to have their posts excluded from analysis. Data were exported from the WhatsApp Messenger group to Microsoft Office Excel 2013 by one of the researchers (PA), who removed identifying information; retained the content, dates, and times of posts; and assigned each member a unique numeric identifier to calculate the percentage of members who wrote posts. All data were then anonymized for qualitative analysis. The messages were analyzed via qualitative content analysis [12]. We approached the data through the lens of the social network theory, which focuses on the effect of social relationships on processing media influence, transferring information, and enabling attitudinal or behavioral change [10]. Two physician researchers (HI and PA) independently coded each post, performed a content analysis of the messages, and produced a list of the common themes that they identified. After this initial review, the researchers discussed their findings, and through discussion, they reached consensus on the themes and created subthemes. The two primary reviewers then independently categorized all posts according to the predetermined themes. Any disagreements were resolved by consensus, and any remaining discordance was brought to the third physician researcher (SAR) and discussed until consensus was achieved. A descriptive quantitative analysis was conducted by using Microsoft Excel 2019 to analyze the frequency of posts within each identified theme. This study was reviewed by the Cleveland Clinic Abu Dhabi Research Ethics Committee and was deemed exempt from institutional review board review, as the data were retrospective, were deidentified, and did not involve any patient information.

### Team Reflexivity

We were cognizant that our research team consisted of 3 female physicians who lived and worked in the Middle East and were frontline workers during the COVID-19 pandemic. To minimize bias, we were blinded to participants' identities. We were mindful of how our experiences influenced our analysis of the data and engaged in frequent group conversations to share, support, and challenge each other's interpretations.

## Results

At the time of data extraction and analysis, there were 161 group members—a 32% increase from the 122 members in the prior year. From February 1, 2020, to May 31, 2020, there were a total of 6101 posts. Of the 161 members, 130 (80.7%) posted at least once during this time period. The number of posts increased 2.6-fold from the number of posts during the same



XSL•FO
RenderX

time period in 2019. Further, 1204 more messages were posted in the chat during this 5-month study interval than during the entire preceding year (6101 posts vs 4897 posts, respectively). Approximately half of all posts (2653/6101, 43.5%) were directly related to the COVID-19 pandemic.

There were 8 general themes identified. Table 1 provides a description of themes and subthemes, example posts, and the relative frequencies of each theme. The most frequent theme was related to requests for information, which represented 46.2% (2818/6101) of all posts. Of the information requests, the majority (1308/2818, 46.4%) were requests for non–COVID-19–related general medical information, and many of these posts consisted of physician referral requests. Of the COVID-19–related information requests, the vast majority (945/1184, 79.8%) were related to logistical information, including quarantine measures, school closures, or mask mandates. Only 3.9% (239/6101) of all posts consisted of specific diagnostic or treatment queries regarding COVID-19.

Approximately 15% (896/6101, 14.7%) of the posts consisted of medical information that was shared with the group by individual members. There were twice as many posts containing information from unverified and non–evidence-based sources (n=517; eg, blogs, social media messages, and local newspaper articles) as there were posts containing evidence-based information (n=297). Several group members expressed confusion and frustration. One group participant noted the following:

> *I have a headache from all the COVID-19 stuff I am reading. I no longer know what to trust and who to believe.*

Another physician stated:

> *This article a perfect example about why a lot of the social media posting of drafts, small series, unproven theories, personal opinion etc. is frankly dangerous. We should defer to only published peer reviewed papers and guidelines. I am personally overwhelmed with all the misinformation I get.*

The frequencies of posts related to each COVID-19 subtheme are displayed in Table 2. In total, 35.6% (945/2653) of COVID-19–related posts were requests for logistical information, and 22.7% (601/2653) of such posts contained supportive or promotional messages related to COVID-19.



XSL•FO
RenderX

**Table 1.** Themes derived from the qualitative analysis of the WhatsApp group chat (posts: N=6101).

| Social network theory principles, general themes, and subthemes | Description of theme | Posts, n (%[a]) | Example posts |
|---|---|---|---|
| **Relationships in the context of general information seeking** | | | |
| **Request for information** | Member requesting information from other group members | 2818 (46.2) | • N/A[b] |
| General (medical) | Information sought on any non–COVID-19–related medical matter | 1308 (21.4) | • "I need a recommendation for a fertility/ IVF center in Abu Dhabi."<br>• "Any pediatric urologist in the group, or one that can be recommended?" |
| COVID-19 logistics | Information sought on logistical topics, such as the location of testing sites, curfew rules, and personal protective equipment protocols | 945 (15.5) | • "Question for the OB in the group: What are your PPE protocol in labor and delivery, patient of unknown COVID status? Any guidelines or protocols on that?"<br>• "Does this mean no school for 4 weeks starting this Sunday?" |
| General (nonmedical) | Information sought on any nonmedical matter | 326 (5.3) | • "Dear ladies, any houseplant experts in the group, particularly orchids."<br>• "Any recommendations for piano repair who will visit your home?" |
| COVID-19 medical | Information sought on COVID-19 diagnoses, symptoms, and treatment protocols | 239 (3.9) | • "Good evening ladies Is there any evidence that fasting and or associated dehydration is a risk favor for worse outcomes in covid19? Just wondering as Ramadan is almost upon us." |
| **Relationships in the context of community building** | | | |
| **Support and promotion** | Member providing moral or emotional support to other group members | 988 (16.2) | • N/A |
| COVID-19 related | Support for COVID-19–related matters | 601 (9.9) | • "Just wanted to thank you all for your support and prayers, my uncle was extubated yesterday, he's officially off the vent but we know he's not out of the woods yet. Hopefully no long-term sequelae from this infection and he'll be able to go back to his family soon." |
| General | Support given for any non–COVID-19–related matters | 387 (6.3) | • "Watch interview with our very own member educating the public on the crucial role anesthesiologists play in delivering safe patient care." |
| Community engagement | Non–medical-related general posts, quotes, memes, videos, or articles of interest | 637 (10.4) | • "I second your gratitude for all the blessings we have...and most importantly each other. Once this is over, I plan to make hugs mandatory amongst all Wonder members" |
| Celebration | Secular and religious holiday wishes, personal and professional milestones, and celebrations | 350 (5.7) | • "Beautiful baby! Wishing him a life full of happiness, health and prosperity." |
| Group administration | Posts for announcing group activities and events as well as for adding and welcoming new group members | 250 (4.1) | • "Reminder: Our Zoom meet is tonight at 730p. We have a dozen WONDER docs signed up already. Anyone else wants to join? DM me your email address. Looking forward to catching up!" |
| Women empowerment | Articles regarding women in medicine, inspirational quotes, and images related to women's empowerment | 73 (1.2) | • "Meet the top 10 Power Businesswomen in the Middle East' ranked by Forbes" |
| **Relationships in the context of information sharing** | | | |
| **Information sharing** | Member sharing information with the group | 896 (14.7) | • N/A |



XSL•FO
RenderX

| Social network theory principles, general themes, and subthemes | Description of theme | Posts, n (%[a]) | Example posts |
|---|---|---|---|
| COVID-19 non–evidence-based source | COVID-19–related information from social media sources | 571 (9.4) | • "…this is from a Romanian physician Facebook group. Seems credible…" |
| COVID-19 evidence-based source | COVID-19–related information from peer-reviewed literature and verifiable sources | 297 (4.9) | • "A great summary interview with Bruce Aylward of the WHO."<br>• "Published today in NEJM: in hospitalized adult patients with severe Covid-19, no benefit was observed with lopinavir–ritonavir." |
| General medical | General medical information from any source | 28 (0.5) | • "You are cordially invited to attend Allergy Connect with Experts." |
| Opportunities | Member seeking or offering employment, volunteer, or donation opportunities | 89 (1.5) | • "Abu Dhabi Blood Bank is running low on reserves as so few people have gone in to donate over the past month. If you know donors, please encourage them to go."<br>• "I have a friend who's a gynecologist in London and looking for a position in Dubai or Abu Dhabi. Please let me know if anyone knows of any vacancies." |

[a]Percentages are rounded to the nearest tenth. Cumulative percentages may not equal 100%.

[b]N/A: not applicable.

**Table 2.** Analysis of the subset of posts related to COVID-19 (n=2653).

| Subtheme | COVID-19–related posts, n (%) |
|---|---|
| Requests for logistical information related to COVID-19 | 945 (35.6) |
| Requests for medical information related to COVID-19 | 239 (9) |
| Sharing of COVID-19–related information from peer-reviewed literature and verifiable sources | 297 (11.2) |
| Sharing of COVID-19–related information from social media and unverifiable sources | 571 (21.5) |
| Supportive, encouraging, or promotional messages related to COVID-19 | 601 (22.7) |

## Discussion

### Use of Social Media

Social media content can provide important insights into the issues and topics that concern health care providers during a global health crisis. Our report on the use of a physician group's WhatsApp chat during the evolving COVID-19 pandemic demonstrates the increased use of this social media forum, and a substantial proportion of the content was related to COVID-19. Social network theory emphasizes the importance of relationships in the context of general information seeking, knowledge sharing, and community building [10]. Accordingly, our study identified 3 main aims of the use of the WhatsApp group, namely information requests, information dissemination, and support and encouragement. A major premise of the social network theory is that a wide network of weaker relationships allows for access to more individuals and resources and can therefore be more beneficial than 1 or 2 strong ties [10]. Throughout the early months of 2020, health care workers cared for large volumes of critically ill patients without any evidence-based therapies while simultaneously dealing with an onslaught of research findings and information [1]. Our findings reflect the confusion and frustration that are often felt by frontline physicians who are trying to navigate a global public health emergency for themselves, their families, and their patients. In our study, group chat members often turned to their colleagues for advice and support. This was evidenced by the large volume of conversations that occurred during this time period and the high engagement levels of members. Compared to the number of posts from the same time period in the previous year, the number of posts increased substantially in 2020. Although there was a modest increase in membership, the exponential rise in the number of posts likely represents the increased use of the platform as a resource for physicians during the extraordinary circumstance of a global pandemic. In fact, 1204 more posts were found during our 5-month study interval than during the entire preceding year (6101 posts vs 4897 posts, respectively). Other studies have reported increased social media use during the COVID-19 pandemic [2,3,6,9].

Almost half (2653/6101, 43.5%) of the posts were directly related to COVID-19, and over one-third (945/2653, 35.6%) of these posts consisted of requests for logistical information. The non–COVID-19–related posts highlighted several important points. For instance, even in the midst of a global pandemic, physicians continued to provide general medical care to their patient populations. Additionally, group members frequently discussed non–COVID-19–related and nonmedical interests. As such, the chat group likely allowed for the opportunity to provide a sense of normalcy to frontline workers in the midst of a public health crisis. It is interesting that the vast majority



XSL•FO
RenderX

of information requests related to COVID-19 dealt with logistics rather than with medical or treatment queries (945/1184, 79.8%). This may have been due to the lack of information available at the time about the novel SARS-CoV-2, but this may also signify the general confusion on public health protocols, which grew as lockdowns, school closures, and social distancing restrictions were implemented. In addition to personally navigating these regulations, physicians were required to guide patients through testing, isolation, and quarantine procedures, often with limited medical knowledge and under frequently changing government policies. Further, as the studied posts were from a women physician group, it should be noted that the group members likely had primary caregiver roles within their families and were personally impacted by the COVID-19 public health protocols. Studies have confirmed that women physicians often bear the majority of childcare and household responsibilities [13,14]. The focus of posts on COVID-19–related logistics may therefore reflect day-to-day priorities that may be shared among female group members. It is notable that community engagement, celebration, and promotion collectively remained important themes during the pandemic. This reflects the tendency of group members to encourage and support each other during uncertain times, which further reinforces the critical role of social media platforms in facilitating a sense of community among medical professionals.

## Study Implications

We are concerned that a substantial majority (571/868, 65.8%) of the posts in the WhatsApp group that contained COVID-19–related information often cited media, social media, or unnamed sources. In fact, only one-third (297/868, 34.2%) of such posts contained information from published medical literature or other verifiable sources. The sharing and discussion of various medical messages within a physician group could serve as a means of critiquing the veracity of information or creating awareness of web-based misinformation. However, the sheer volume of unverified posts and the frequent deviations from the evidence-based data that physicians are expected to disseminate could result in confusion and the inadvertent spread of misinformation from physicians to their patients, as evidenced by several chat participants who expressed concern and frustration about the posting of unverified information by other group members. This phenomenon is not unique to this one chat group. A previous study on the prevalence of misinformation in tweets about health care found that approximately 20% of tweets were inaccurate [15]. In fact, one of the most widely spread conspiracy theories, which linked COVID-19 to the 5G network, was traced to comments made by a Belgian physician in January 2020 [16]. The director-general of the World Health Organization described the large volume of unproven or inaccurate information on social media during the pandemic as an *infodemic of misinformation* [17]. The confusion resulting from misinformation can ultimately cause physicians to question the legitimacy of new scientific discoveries regarding effective COVID-19 therapeutics or vaccines. Moreover, the high volume of social media posts can result in information fatigue. Despite the understandable desire of physicians to share and receive useful information, without verifying the veracity and credibility of information prior to sharing it, physicians risk unwittingly facilitating the dissemination of misinformation. This may reflect a lack of formal social media education among physicians [18], which is compounded by the considerable challenges of analyzing data during an evolving pandemic [19]. More and more social interactions are occurring on web-based platforms. As such, physicians should be more cognizant of appropriate and effective social media use in the context of data analysis, synthesis, and sharing. Our findings have several important implications. First, the results substantiate the critical role that social media platforms play in facilitating communication and fostering connectedness among physicians coping with population health crises. Second, our study also provides insights on content and topics that seem to be the most relevant to physician communities during such crises. Lastly, our data reflect areas of concern regarding the use of social media in these professional communities during times of uncertainty and can be used to inform the design of future interventions and research.

In addition to confirming the results of prior work, this study highlights the need for additional research into the evidence-based approaches that physicians use to analyze health information obtained from social media. Similar concerns have been raised about the increasing dissemination of medical information through social media, including the lack of editorial oversight for web-based data and the harm caused by the rapid dissemination of incorrect medical information [20].

## Limitations

Although this study involves a single international physician group, the findings likely reflect the common challenges faced by health care workers who deliver health services in dynamic logistical and biomedical environments that are intensified by global health crises. The study group contained only women physicians; therefore, our findings may represent priorities that correlate with gendered experiences and roles. To our knowledge, the existing literature lacks published studies that report on gender differences in physicians' social media behaviors, though studies have suggested that there are gender differences in success on social media platforms [21,22]. Lastly, it is unknown whether group messages actually impacted physicians' attitudes or behaviors.

## Conclusion

The uncertainties posed by an evolving global health crisis represent considerable challenges to the health care workforce. As the world has become increasingly more connected through social media, these platforms represent critical information dissemination tools. Our findings confirmed the importance of social media in creating a communicative and collaborative platform for physicians in the midst of a public health emergency. Although more accessible information can undoubtedly benefit patient care, our findings raised important questions regarding the accuracy and credibility of shared information. Larger multinational infoveillance studies are needed to better understand social media discourse among physicians during public health crises.



XSL•FO
RenderX

## Authors' Contributions

SAR led the drafting of the manuscript. All authors contributed equally to the conception, analysis, editing, and final approval of the manuscript. All authors had equal access to the data and had final responsibility for the publication of this study.

## Conflicts of Interest

None declared.

## References


1. da Silva JAT, Tsigaris P, Erfanmanesh M. Publishing volumes in major databases related to Covid-19. Scientometrics 2020 Aug 28:1-12 [FREE Full text] [doi: 10.1007/s11192-020-03675-3] [Medline: 32904414]
2. Park HW, Park S, Chong M. Conversations and medical news frames on Twitter: Infodemiological study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
3. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
4. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: A systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
5. Steffens MS, Dunn AG, Wiley KE, Leask J. How organisations promoting vaccination respond to misinformation on social media: a qualitative investigation. BMC Public Health 2019 Oct 23;19(1):1348 [FREE Full text] [doi: 10.1186/s12889-019-7659-3] [Medline: 31640660]
6. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: Retrospective observational infoveillance study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]
7. COVID-19 USA Physician/APP group. Facebook. URL: https://www.facebook.com/groups/815998962228639/ [accessed 2020-12-30]
8. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining physicians' opinions on social media to obtain insights into COVID-19: Mixed methods analysis. JMIR Public Health Surveill 2020 Jun 18;6(2):e19276 [FREE Full text] [doi: 10.2196/19276] [Medline: 32421686]
9. Ibrahim H, Anglade P, Abdel-Razig S. The use of social media by female physicians in an international setting: A mixed methods study of a group WhatsApp chat. Womens Health Rep (New Rochelle) 2020 Feb 17;1(1):60-64 [FREE Full text] [doi: 10.1089/whr.2019.0015] [Medline: 33786474]
10. Brass DJ. A social network perspective on organizational psychology. In: Kozlowski SWJ, editor. The Oxford Handbook of Organizational Psychology, Volume 1. New York, New York, US: Oxford University Press; Jun 2012.
11. United Arab Emirates: WHO coronavirus disease (COVID-19) dashboard with vaccination data. World Health Organization. URL: https://covid19.who.int/region/emro/country/ae [accessed 2020-12-30]
12. Bloor M, Wood F. Keywords in Qualitative Methods: A Vocabulary of Research Concepts. Thousand Oaks, California: Sage Publications Ltd; 2006.
13. Buddeberg-Fischer B, Stamm M, Buddeberg C, Bauer G, Häemmig O, Knecht M, et al. The impact of gender and parenthood on physicians' careers--professional and personal situation seven years after graduation. BMC Health Serv Res 2010 Feb 18;10:40 [FREE Full text] [doi: 10.1186/1472-6963-10-40] [Medline: 20167075]
14. ALobaid AM, Gosling CM, Khasawneh E, McKenna L, Williams B. Challenges faced by female healthcare professionals in the workforce: A scoping review. J Multidiscip Healthc 2020 Aug 05;13:681-691 [FREE Full text] [doi: 10.2147/JMDH.S254922] [Medline: 32821112]
15. Chretien KC, Azar J, Kind T. Physicians on Twitter. JAMA 2011 Feb 09;305(6):566-568. [doi: 10.1001/jama.2011.68] [Medline: 21304081]
16. Shanapinda S. No, 5G radiation doesn't cause or spread the coronavirus. Saying it does is destructive. The Conversation. 2020 Apr 07. URL: https://theconversation.com/no-5g-radiation-doesnt-cause-or-spread-the-coronavirus-saying-it-does-is-destructive-135695 [accessed 2020-12-30]
17. Countries urged to act against COVID-19 'infodemic'. UN News. 2020 Sep 23. URL: https://news.un.org/en/story/2020/09/1073302 [accessed 2020-12-30]
18. Gagnon K, Sabus C. Professionalism in a digital age: opportunities and considerations for using social media in health care. Phys Ther 2015 Mar;95(3):406-414. [doi: 10.2522/ptj.20130227] [Medline: 24903111]
19. Wolkewitz M, Puljak L. Methodological challenges of analysing COVID-19 data during the pandemic. BMC Med Res Methodol 2020 Apr 14;20(1):81 [FREE Full text] [doi: 10.1186/s12874-020-00972-6] [Medline: 32290816]
20. Chan T, Trueger NS, Roland D, Thoma B. Evidence-based medicine in the era of social media: Scholarly engagement through participation and online interaction. CJEM 2018 Jan;20(1):3-8. [doi: 10.1017/cem.2016.407] [Medline: 28077195]




XSL•FO
RenderX

21. Woitowich NC, Arora VM, Pendergrast T, Gottlieb M, Trueger NS, Jain S. Gender differences in physician use of social media for professional advancement. JAMA Netw Open 2021 May 03;4(5):e219834 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.9834] [Medline: 33983403]
22. Shillcutt SK, Silver JK. Social media and advancement of women physicians. N Engl J Med 2018 Jun 14;378(24):2342-2345. [doi: 10.1056/NEJMms1801980] [Medline: 29897857]

## Abbreviations

**UAE:** United Arab Emirates
**WONDER:** Women Doctors in the Emirates

*Edited by G Eysenbach; submitted 05.07.21; peer-reviewed by M Salimi, P van der Keylen; comments to author 30.07.21; revised version received 27.08.21; accepted 18.09.21; published 07.12.21.*

*Please cite as:*
*Abdel-Razig S, Anglade P, Ibrahim H*
*Impact of the COVID-19 Pandemic on a Physician Group's WhatsApp Chat: Qualitative Content Analysis*
*JMIR Form Res 2021;5(12):e31791*
*URL: https://formative.jmir.org/2021/12/e31791*
*doi:10.2196/31791*
*PMID:34784291*

©Sawsan Abdel-Razig, Pascale Anglade, Halah Ibrahim. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Nursing Perspectives on the Impacts of COVID-19: Social Media Content Analysis

Ainat Koren[1*], PhD, DNP; Mohammad Arif Ul Alam[2*], PhD; Sravani Koneru[2], MSc; Alexa DeVito[1]; Lisa Abdallah[1], PhD; Benyuan Liu[2], PhD

[1]Solomont School of Nursing, University of Massachusetts Lowell, Lowell, MA, United States

[2]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States

[*]these authors contributed equally

**Corresponding Author:**
Ainat Koren, PhD, DNP
Solomont School of Nursing
University of Massachusetts Lowell
113 Wilder Street, Suite 200
Health and Social Science Building
Lowell, MA, 01854-3058
United States
Phone: 1 9789344429
Email: Ainat_Koren@uml.edu

## Abstract

**Background:** Nurses are at the forefront of the COVID-19 pandemic. During the pandemic, nurses have faced an elevated risk of exposure and have experienced the hazards related to a novel virus. While being heralded as lifesaving heroes on the front lines of the pandemic, nurses have experienced more physical, mental, and psychosocial problems as a consequence of the COVID-19 outbreak. Social media discussions by nursing professionals participating in publicly formed Facebook groups constitute a valuable resource that offers longitudinal insights.

**Objective:** This study aimed to explore how COVID-19 impacted nurses through capturing public sentiments expressed by nurses on a social media discussion platform and how these sentiments changed over time.

**Methods:** We collected over 110,993 Facebook discussion posts and comments in an open COVID-19 group for nurses from March 2020 until the end of November 2020. Scraping of deidentified offline HTML tags on social media posts and comments was performed. Using subject-matter expert opinions and social media analytics (ie, topic modeling, information retrieval, and sentiment analysis), we performed a human-in-a-loop analysis of nursing professionals' key perspectives to identify trends of the COVID-19 impact among at-risk nursing communities. We further investigated the key insights of the trends of the nursing professionals' perspectives by detecting temporal changes of comments related to emotional effects, feelings of frustration, impacts of isolation, shortage of safety equipment, and frequency of safety equipment uses. Anonymous quotes were highlighted to add context to the data.

**Results:** We determined that COVID-19 impacted nurses' physical, mental, and psychosocial health as expressed in the form of emotional distress, anger, anxiety, frustration, loneliness, and isolation. Major topics discussed by nurses were related to work during a pandemic, misinformation spread by the media, improper personal protective equipment (PPE), PPE side effects, the effects of testing positive for COVID-19, and lost days of work related to illness.

**Conclusions:** Public Facebook nursing groups are venues for nurses to express their experiences, opinions, and concerns and can offer researchers an important insight into understanding the COVID-19 impact on health care workers.

***(JMIR Form Res 2021;5(12):e31358)*** doi:10.2196/31358

**KEYWORDS**

mental health; information retrieval; coronavirus; COVID-19; nursing; nurses; health care workers; pandemic; impact; social media analytics



XSL•FO
RenderX

## Introduction

### Background

Nursing is an occupation with unique potential for exposure to environmental and occupational hazards in the work setting. Nurses confront potential exposure to infectious diseases, toxic substances, stress, back injuries, and radiation [1]. The COVID-19 epidemic poses a unique, health risk situation that is rapidly evolving [2]. The American Nurses Association Code of Ethics (2015) states that the nursing profession's nonnegotiable ethical practice standard, according to Provision 2 of the Code, is that "the nurse's primary commitment is to the patient." Provision 5 of the Code states that "the nurse owes the same duty to protect themselves" [3]. These two equal obligations can be in conflict during pandemics, when nurses must continually care for critically ill infectious patients under extreme circumstances, including insufficient or inadequate resources and uncontained contagious diseases.

Professional nurses historically bring compassionate competent care to disaster responses but are faced with challenges to provide care when the nature of their work puts them at increased risk. Nurses struggle with feeling physically unsafe in the disaster response situation, such as in times of scarce resources where supplies of personal protective equipment (PPE) may be inadequate [4]. Nurses are concerned about professional, ethical, and legal protection when asked to provide care in such high-risk situations, such as the COVID-19 pandemic. According to DeWolfe, disasters such as the COVID-19 pandemic impact those who experience them psychologically and socially [5]. Whether one considers the COVID-19 pandemic a human-caused or natural disaster, the human effects of living through such an experience are significant, especially when exposure to such a disaster is felt personally. For example, nurses working on the front lines throughout the COVID-19 pandemic have felt a direct effect of this disaster and, therefore, could experience an unusually large number of psychological and social reactions to this experience [5]. DeWolfe explains that high-exposure survivors, such as nurses and other health care workers, could experience a range of effects, such as anxiety, depression, sadness, posttrauma symptoms, somatic symptoms, and substance abuse [5].

Researchers around the world who have been examining the psychological impact on nurses and other health care workers as a result of the COVID-19 pandemic have shown that nurses and other health care workers are experiencing high anxiety and fear, especially as these relate to concerns of infecting family members, being unable to socialize, and transmission of COVID-19 in their work settings [6]. A cross-sectional descriptive analysis of 204 COVID-19–infected health care workers showed that not only does the lack of PPE put nurses at risk of contracting COVID-19 from patients, but the lack of compliance from fellow employees to wear masks and practice social distancing, especially during breaks, puts nurses at risk [6]. An et al found that depressive symptoms among emergency department nurses in China were common, and those reporting higher depressive symptoms also reported lower quality of life [7].

Hu et al examined frontline nurses in Wuhan, China; their findings demonstrated that nurses experienced moderate to high levels of anxiety, depression, burnout, and fear, along with reporting having one or more skin lesions [8,9]. Nurses are also facing ethical dilemmas, such as which patients to prioritize and who should receive ventilation because of a lack of a sufficient number of ventilators [10] as well as moral distress related to uncertainty about their skills to tackle the virus [10]. Qualitative studies have demonstrated that nurses in China who were providing direct care to patients with COVID-19 experienced a range of both positive and negative emotions [11]. Liu et al identified key themes that stress the emotional toll being experienced by nurses, specifically related to feelings of facing challenges and danger, fear of being infected, exhaustion, and stress [12]. Along with these feelings, nurses also expressed their strong sense of duty and responsibility for being a health care provider during this pandemic, along with the hope that the epidemic would be overcome [12].

Sun et al found that in the early weeks of dealing with the pandemic, nurses primarily experienced negative emotions, such as fatigue, discomfort, helplessness, fear, and anxiety [11]; however, with time working in the setting and with knowledge growth of the care they provided, nurses expressed many positive emotions, such as those focusing on coping and self-care, confidence in their self-prevention of contracting COVID-19, and happiness gained from their patients' respect and from their family and team support. The stress of working with patients with COVID-19 carries over into the daily life of nurses, as they feel isolated from family and friends as well as having their children's caregivers quit because of fear of infection and being unable to attend funerals of loved ones [6]. Some nurses became frustrated as they found themselves out of work for the first time and wished they could do more to help [10].

### Gap in Knowledge

Various forms of media have played a major role in the COVID-19 pandemic, being major sources of information for the public. However, these media sources have presented contradictory opinions and viewpoints about the virus, causing some to take the virus less seriously than others, leading to more distress for nurses [10]. The science of understanding health-related information that is distributed via social media to inform public health and public policy, known as infodemiology, has been particularly useful for identifying disease outbreak patterns and studying public perceptions of various diseases. Analysis of health event data posted on social media platforms not only provides firsthand evidence of health event occurrences but also enables faster access to real-time information that can help health professionals and policy makers frame appropriate responses to health-related events. Nurses have begun to use social media as a voice for health care workers on the front lines. Online videos have surfaced, showing the chaos of hospital wings, and firsthand accounts of the traumas and struggles nurses have faced have appeared on sites such as Facebook [10,13].

Nurses are using social media in order to communicate with the public and advocate for more supplies and support [13]. For



XSL•FO
RenderX

example, the "#GetMePPE" hashtag on Twitter was generated in order to spread awareness of the PPE shortages; this led to creation of a petition with over 62,000 signatories, which combined with a website, GetUsPPE.org, to allow health care workers to obtain hundreds of thousands of articles of PPE [14]. The COVID-19 outbreak has resulted in a set of studies that have examined public perceptions, thoughts, and concerns about this pandemic through the use of social media data. All of these studies relied on data from public digital media, such as Twitter or Weibo platforms; these studies analyzed data from early periods of the pandemic, using different sentiment analysis techniques on the general population, irrespective of users' professions, or evolution of sentiments over time on temporal events. In this study, we specifically analyzed social media discussions by nursing professionals participating in publicly formed Facebook groups to develop longitudinal insights related to the pandemic's impacts in terms of what health care providers experienced over time.

### Study Aims

The primary aim of this study was to explore nurses' work experiences in dealing with the coronavirus pandemic and how it affected their emotional state. To achieve this, we specifically employed sentiment analysis, topic modeling, and information retrieval techniques to estimate the influence of physical, mental, and psychosocial factors of nurses related to the COVID-19 pandemic. The analysis captured major themes presented by the nurses who participated in publicly available social media groups from March to the end of November 2020. The analysis examined comments made to the posts as well. Specifically, we analyzed the major topics of concern that were posted by nurses (eg, lack of masks, PPE, and ventilators; fear of being infected; family difficulty; and worrying about employment). The major topics were identified, guided by findings presented in recent publications. The analysis also focused on how these topics changed over time (eg, from medical equipment shortages at the beginning of the pandemic to treatment in later stages). In addition, using a sentiment analysis technique, we analyzed the feelings and emotions, both positive and negative, expressed in the posts and comments.

This study was reviewed by the University of Massachusetts Lowell Institutional Review Board (IRB) and was determined to be exempt from review.

### Social Media Analytics State of the Art

Various approaches were used for text sentiment extraction by researchers, which can be divided into four categories: keyword, lexicon, machine learning, and hybrid. Some researchers also used linguistic rule-based methods [15], keyword-based methods [16], emotion-based models [17,18], natural language processing (NLP) [19], and case-based reasoning [20]. Keyword-based methods detect sentiment by looking for a match between words in a piece of text and emotion keywords, providing a matching index, which is also called information retrieval [16]. Lexicon-based methods use a sentiment lexicon or dictionary to detect the correct emotion from a piece of text [21]. Machine learning methods use both supervised [22,23] and unsupervised [24,25] learning for emotion detection, using various existing classification and clustering methods. Hybrid methods merge more than one of the above techniques and apply the results to recognize text emotion [16,26-28]. Emotion is generally defined and described by various emotion models. All existing emotion models can be divided into categorical and dimensional models [29]. Categorical emotion models, such as those by Ekman [30], Shaver [31], and Oatley [32], categorize all human emotions into a few major classes (eg, anger, disgust, fear, joy, love, positive, and negative). In contrast, dimensional sentiment models, such as Plutchik's model [33], the circumplex model [34], the cognitive structure of emotions model [35], and Loveim's model [34], classify sentiment in detail, using multiple dimensions (ie, valence, arousal, and dominance) and intensities (ie, basic, mild, and intense) in a question-and-answer form. We used the most popular methods from the existing literature—the information retrieval technique (keyword), predefined dictionary-based Linguistic Inquiry and Word Count (lexicon), and pretrained Bidirectional Encoder Representations from Transformers (BERT; machine learning) [36]—to identify sentiments in various use cases.

## Methods

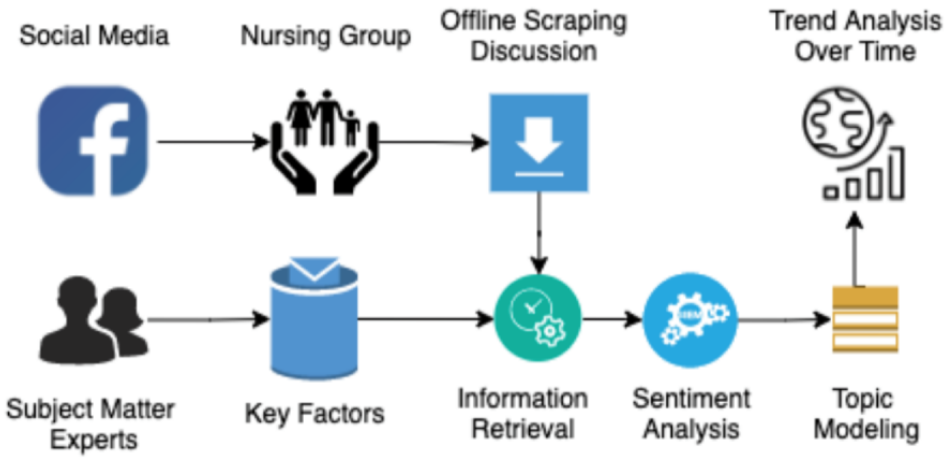
### Overview

Social media refers to digital platforms where people can express their ideas, providing easy access to a diverse population all over the world. In particular, as of November 2020, Facebook, with 2.7 billion monthly active users, is the largest platform and plays a dominant role in social networks. In this study, we applied data mining techniques with added quotes to understand nursing professionals' perspectives regarding the COVID-19 pandemic as discussed in trusted open Facebook groups of nurses. Figure 1 illustrates the flowchart of our methodology.



XSL•FO
RenderX

**Figure 1.** Methodology flowchart.




## Data Preparation

Our data preparation phase consisted of group selection, data collection, and preprocessing. First, we selected a public nursing professionals' group, which consisted of 108,354 members, formed by nursing professionals with the sole purpose of COVID-19–related discussion, and we collected the nurses' posts. Data collection from Facebook group posts is more challenging than the use of any other similar social media platforms, such as Twitter. Since Facebook's application programming interface (API) lacks the capability to extract comments and other necessary information (eg, reactions and photos), we used the Facebook HTML page offline downloader and parsed the HTML tags using the Beautiful Soup library from Python (Python Software Foundation) to extract the following information: various post IDs, hash value of usernames (deidentified), post text, number of likes, date, and the comments for each post. To represent the emotion pattern during the pandemic situation, we collected posts from the beginning of the pandemic, on March 1, 2020, until November 30, 2020, and saved them in a relational database of two tables—main posts and related comments—with appropriate private-public keys definitions. The collected raw data contained background noise, such as URLs, hashtags, emoji, stop words, and empty posts, which was removed from the data using Python-based data cleaning tools in order to provide increased precision scores.

## Data Analysis Tools

We used two different analytical tools to analyze collected data: the sentiment analysis tool and the information retrieval tool.

### *Sentiment Analysis Tool*

Sentiment analysis is used as a process to determine the character of a text (ie, positive, negative, or neutral), assisting one to understand overall perceptions regarding a topic of conversation. BERT is a transformer-based machine learning technique for NLP pretraining developed by Google to extract sentiments. We initially trained and validated our BERT-based supervised model on an existing Twitter data set of 1.6 million items [37]. The Twitter data set has four labels: joy, sadness, anger, and fear. For this research study, we used the BERT [36] framework to extract sentiments from the selected data texts.

### *Information Retrieval Tool*

Information retrieval is a process of getting information or phrases out of the document repository. More specifically, the information retrieval tool returns texts from the database that consist of the information queried by users in the form of texts, sentences, or phrases, in order to represent top ranking or similarity scores. For this research study, we used the Python-based information retrieval tool Whoosh (Anserini), which can take either phrases, words, or documents of text or a set of conditional phrases, connected with the "and/or" relation, and return related posts of existence of queried phrases with confidence (Multimedia Appendix 1) [38].

## Validity of Choosing Analytic Tools

### *Overview*

To analyze the nursing professionals' perspectives of the COVID-19 outbreak (ie, nurses' psychology [39,40], decision making [41,42], emotions [43], and concerns [44]), we applied current social media text analytic techniques [44,45]. In this section, we explain the validity of selecting BERT for sentiment analysis and Whoosh for information retrieval tools.

### *Bidirectional Encoder Representations From Transformers*

NLP is one of the most cumbersome types of machine learning methods in the area of data preprocessing. Apart from the preprocessing and tokenizing of text data sets, it takes a great deal of time to train successful NLP models. In 2018, a team of Google scientists proposed and open-sourced BERT, a major breakthrough that took the deep learning community by storm because of its incredible performance. BERT is a transformer-based machine learning technique for NLP pretraining methods [36,46]. As per a Google scholarly citation, which has been cited over 21,000 times, BERT has been considered the most popular sentiment analysis tool for use with social media posts. There are two pretrained general BERT variations: (1) BERT-Base—a 12-layer, 768-hidden, 12-head, 110-million–parameter neural network architecture, and (2) BERT-Large—a 24-layer, 1024-hidden, 16-head, 340-million–parameter neural network architecture. Both of the BERT models have been trained on English Wikipedia (2500 million words) and BookCorpus (800 million words) and



XSL•FO
RenderX

achieved the best accuracies for some of the NLP tasks, such as sentiment analysis [47,48]. In this paper, we used a pretrained BERT model proposed by Dai et al for extracting sentiments from social media posts [49]. This particular model used the original vocabulary of BERT-Base as its underlying word piece vocabulary and used the pretrained weights from the original BERT-Base as the initialization weights. Then, the model used English tweets from September 1 to October 30, 2018, to pretrain the BERT-Base model on a total of 60 million English tweets, consisting of 0.9 billion tokens. This particular BERT model achieved remarkable accuracies on sentiment analysis (>91% accuracy on Twitter posts) and fake news detection (>98% accuracy on Twitter posts), which inspired us to choose this pretrained model for our study [49] (Table 1).

**Table 1.** Performance of our pretrained BERT model compared with another model.

| Target text type | BERT[a]-Base model, % | Pretrained BERT model on target, % |
|---|---|---|
| **Tweets** | | |
| Precision | 89.9 | 91.7 |
| Recall | 89.4 | 91.1 |
| F1 score | 88.0 | 89.5 |
| **Forum posts** | | |
| Precision | 92.6 | 93.8 |
| Recall | 92.4 | 93.4 |
| F1 score | 92.2 | 93.0 |

[a]BERT: Bidirectional Encoder Representations from Transformers.

### *Anserini Tool*

Anserini is an open-source software toolkit for the Lucene-based search engine via information retrieval toward building real-world search applications [38]. The Lucene-based search engine (Apache Lucene), first proposed in a Lucene4IR paper [50] and later improved by Grand et al [51] and Kamphuis et al [52], is widely used and is a standard foundation for search applications. The central purpose of the Anserini engine is to provide ranking (ie, indexes) of documents and sentences based on searched expression. The core components of the Anserini architecture are a multi-threaded indexing engine or wrapper, a streamlined information retrieval evaluator, and a relevance feedback engine. The wrapper provides abstractions for document collections as well as implementing an efficient, high-throughput, and multi-threaded indexer that takes advantage of these abstractions. The evaluator develops a multistage ranking architecture by extracting document features from the abstraction. The feedback component develops a relevance feedback index based on a vocabulary mismatch method between searched expressions and document collections. The final output index represents the ranking of similarity index values, where a higher value means greater similarity. We used Anserini for identifying the existence of COVID-19–related key information from the social media posts [38].

### Data Analysis

We used state-of-the-art NLP for cleaning, topic modeling, sentiment analysis, and information retrieval. During the data cleaning step, we removed background noise, such as URLs, hashtags, emoji, stop words, and empty posts, from the entire data set to increase the precision score. Then, we used BERT for detecting sentiments. In this process, we used Hugging Face's transformer library written in TensorFlow to label our collected data with sentiments, along with the frequency [53]. Hugging Face is a Python-based transformer library that can support our pretrained BERT model and can be used to label any collected data with sentiments. This library specifically shows the potential sentiments from a text, and one needs to confirm the sentiments from an interface. One graduate student was engaged to confirm the sentiments from the Hugging Face interface. It should be mentioned that one or more sentiments can be associated with a single post; thus, a single post can be associated with multiple sentiments. In that case, a single post can be considered multiple times for multiple sentiments. These detected sentiments were used and subdivided into subtopics later. After getting the emotions measure, we explored additional topics (eg, lack of masks, PPE, and ventilators; fear of being infected; family difficulty; and worrying about employment), which are specific and cannot be detected or retrieved by use of sentiment analysis or topic modeling methods. Therefore, we used an information retrieval technique (Anserini) to further label posts. Anserini is a Python-based search engine, similar to Lucene search indexing. This will result in the posts on these topics along with the score, which is the term frequency–inverse document frequency for the topic [38]. Based on the emotional themes identified, specific anonymized posts and comments were included in this paper to highlight qualitative examples of nurses' own words.

## Results

### Overview

The following results illuminate the negative and positive emotions expressed by nurses over time. The emotions are related to a variety of identified nurses' experiences during a 9-month period (ie, March 1 to November 30, 2020) of the COVID-19 pandemic. Sample data (ie, comments and posts) are displayed in Table 2.



XSL•FO
RenderX

**Table 2.** Distribution of posts and comments over 9 months in 2020.

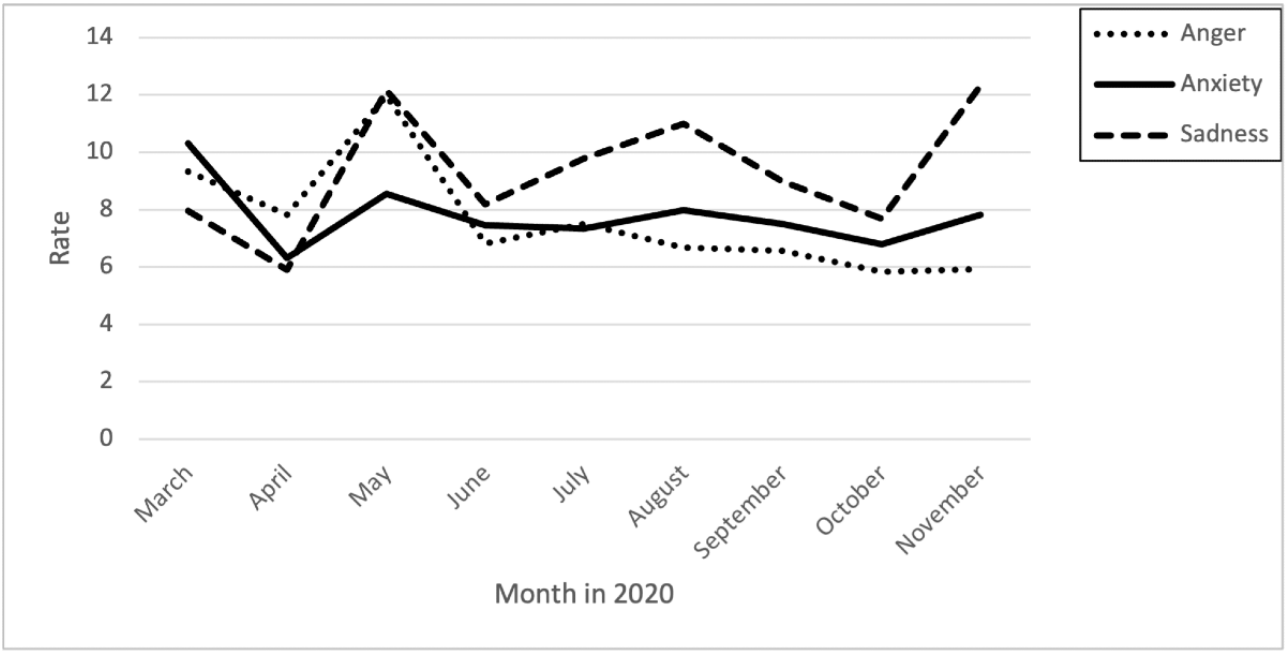
| Month | Posts[a] (n=1548), n (%) | Comments[a] (n=109,445), n (%) |
|---|---|---|
| March | 8 (0.5) | 1739 (1.6) |
| April | 7 (0.5) | 1939 (1.8) |
| May | 64 (4.1) | 11,432 (10.4) |
| June | 111 (7.2) | 11,777 (10.8) |
| July | 144 (9.1) | 16,627 (15.2) |
| August | 457 (29.5) | 16,553 (15.1) |
| September | 218 (14.1) | 24,274 (22.2) |
| October | 178 (11.5) | 8313 (7.6) |
| November | 361 (23.3) | 16,791 (15.3) |

[a]There were a total of 110,993 posts and comments combined.

## Detection of Negative Emotions Expressed by Nurses Over Time: Anger, Anxiety, and Sadness

Figure 2 shows the rate variation among the emotions of sadness, anger, and anxiety. The rate was calculated by dividing the number of specific posts and comments of the expressed emotion by the overall emotional posts and comments for the month. The displayed trend demonstrates that rates of all emotions (ie, posts and comments) peaked in May, July, and August. Sadness and anxiety rates showed an additional peak in November, while the rate of anger was trending down (Figure 2).

**Figure 2.** Rates of posts and comments related to the emotions of anger, anxiety, and sadness over time. The rate was calculated by dividing the number of specific posts and comments of the expressed emotion by the overall emotional posts and comments for the month.




Sample posts and comments that exemplify these emotions are shared in this section. One nurse posted the following comment that displays anxiety with her role during the pandemic:

> *I am terrified we will end up hospitalized or dead. My chest feels tight, but I think it's anxiety and not a COVID symptom.*

One nurse posted a comment that demonstrates anger related to undefined policies on returning to work after exposure to COVID-19:

> *I am so mad they let a nurse who is COVID positive back to work with no rescreening done. He gets two hours into shift and has to go home sick. Thanks for the exposure people*

One nurse shared her feelings about depression and unhappiness, which led to reconsidering her profession as a result:

> *...This pandemic is absolutely draining and has even made me reconsider nursing. I am currently making a slight change and jumping into resource nursing. I've worked the COVID ICU [intensive care unit] now for months and have noticed myself progressively becoming more depressed and unhappy. I'm making this change for mine and my family's sanity...*

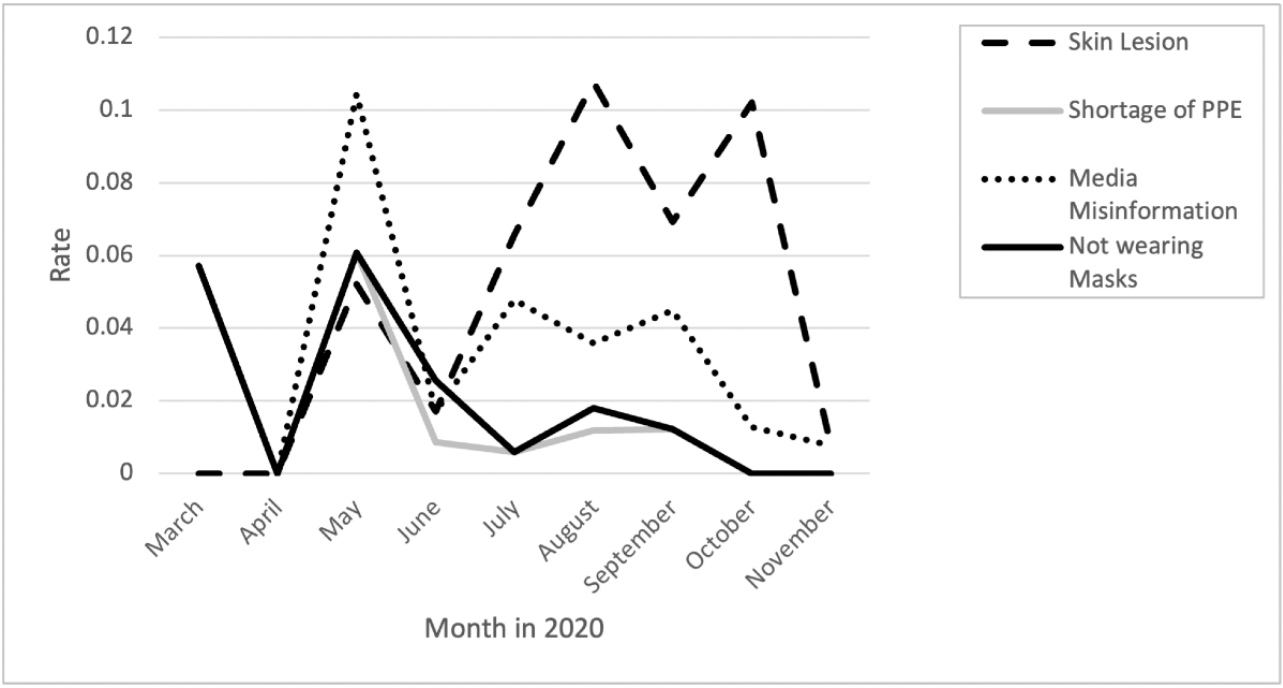


XSL•FO
RenderX

### Frustration Due to Mask Side Effects, Shortage of PPE, Media Misinformation, and Lack of Compliance With Masks

Frustration with misinformation from the media was ongoing, with the highest peaks from April through June. Frustration caused by shortage of PPE peaked from April through June. Frustration from skin lesions was ongoing, with the highest peaks in August and October. Frustration due to people not complying with mask recommendations peaked from April through July and again from July through September (Figure 3).

**Figure 3.** Rates of posts and comments related to personal protective equipment (PPE) and misinformation from media over time.




The following posts and comments illuminated the emotions expressed as they relate to frustration related to skin lesions, shortage of PPE, misinformation from media, and people not complying with mask recommendations.

Nurses posted comments about their struggle with mask rashes and lesions caused by wearing masks all day:

> *I'm starting to get pressure ulcer on the tip of my nose from 12 hr shifts w surgical masks on...*

Nurses also reported receiving improper PPE:

> *They gave us surgical masks and then when COVID probable, they didn't give us N95s until cases were exponentially increasing. FYI (I had my own N95 and brought it). Then they gave us these N95s that were not fitted - that broke when I was using it (I had to staple the straps)*

Another nurse shared that her facility controlled their access to PPE:

> *Heck, our acute care hospital locks up the PPE. We have to sign it out and are only allowed 1 mask a shift*

Some nurses posted comments about frustration with those who refuse to wear masks at work:

> *Anyone else returning from work after being sick with COVID annoyed/anxious over staff removing their mask at the nurses' station all day long? Worried for my staff getting sick and tired of people just not caring. I get it we are all sooo tired of this but COVID is still here.*

A few nurses expressed their concerns about the spread of misinformation about the virus:

> *...I'm so sick and tired of people with ZERO credentials and experience in the medical field telling others the virus is a hoax and wearing a mask is pointless and literally trying to convince others this virus isn't a problem. People are so shortsighted on their little soapbox that they don't realize PEOPLE are DYING and their constant ramming of conspiracy theories down people's throats could be enough to convince someone this virus isn't deadly and can get someone killed. I'm so irritated right now.*

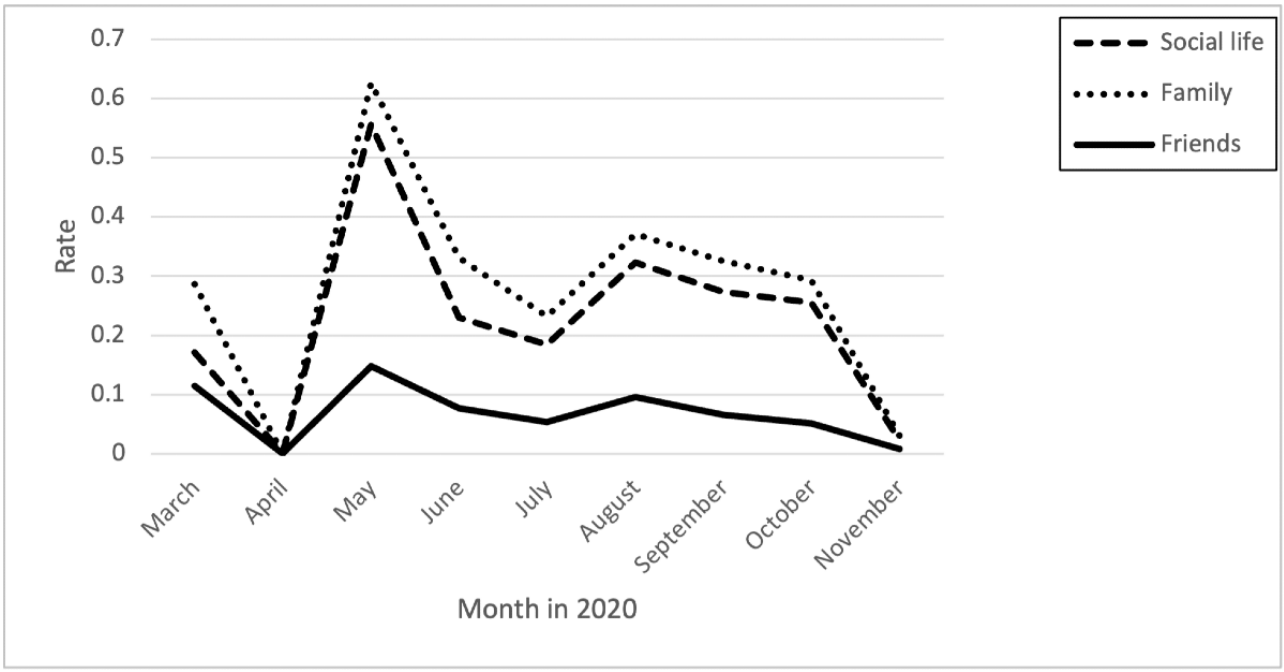
### Isolation as it Relates to Social Life, Family, and Friends

The rate of isolation-related posts and comments across all categories peaked across all months, with the highest rates from April through June, followed by another rate increase in July to October (Figure 4). Nurses expressed concerns about isolation from family due to fear of infecting them:



XSL•FO
RenderX

**Figure 4.** Rates of posts and comments related to isolation over time.



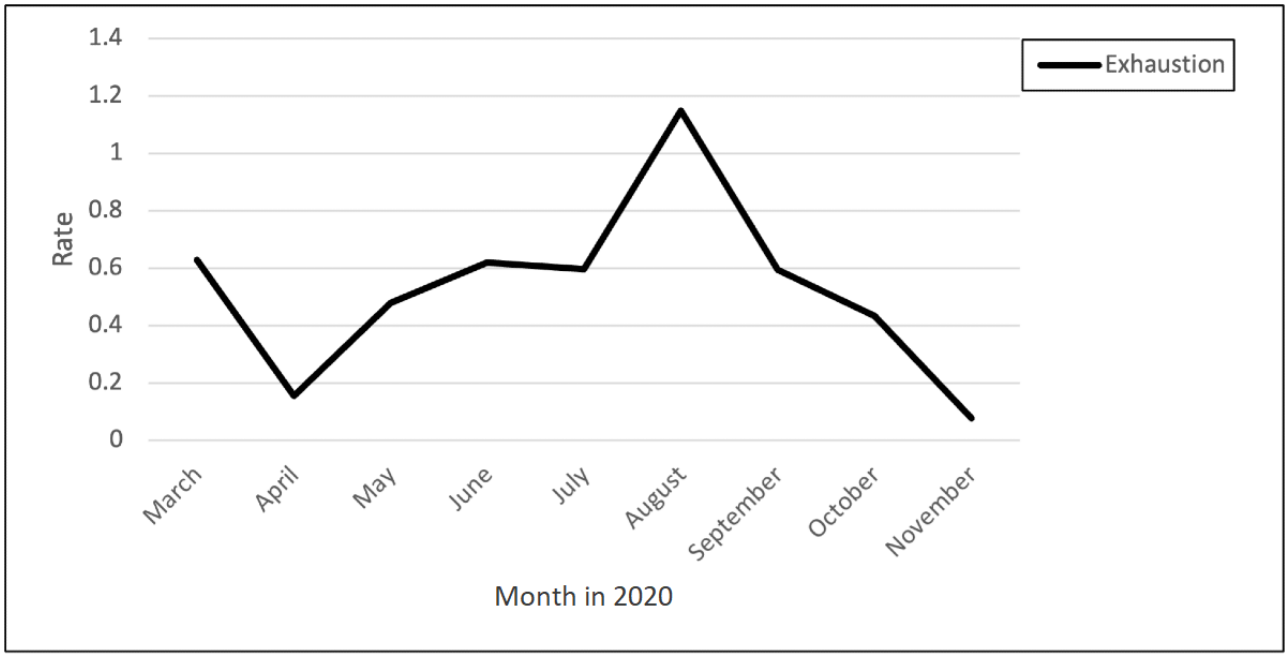

*I don't personally care about the risk to myself, it's more the fact that I'd like to be able to see my parents again and possibly hug my mom at least once this year. She has 3 autoimmune diseases. Me being in the same room as her is a major risk.*

*I miss my kids. I won't go near them. haven't in 3 weeks. I talk to them from 10-12 feet away with a mask. It sucks.*

## Exhaustion and Loneliness

The displayed trend demonstrates that exhaustion peaked over time, with a significant peak from July through September (Figures 5 and 6). Nurses described the mental and emotional exhaustion of watching patients decline from COVID-19:

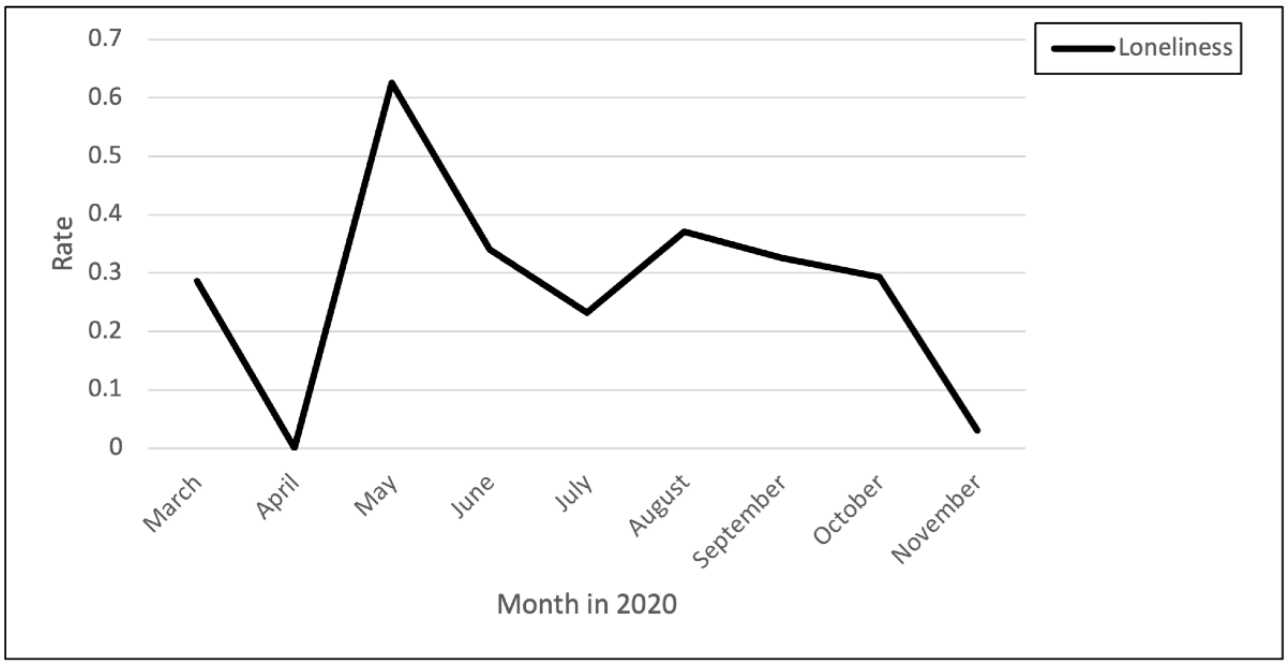
**Figure 5.** Rates of posts and comments related to exhaustion over time.






XSL•FO
RenderX

**Figure 6.** Rates of posts and comments related to loneliness over time.




> *Seeing these people suddenly tank and say goodbye before they get put on the vent. Face timing their loved ones, one last time, it don't get any easier. How do you fit a lifetime of love and relationships into a 2-minute phone call? I am struggling with asking to take a break from my unit. I am exhausted mentally and emotionally.*

The displayed trend demonstrates that loneliness peaked from April to the end of May, with another peak from July through October. One nurse commented on how the isolation and longing to hug their family has led them to question their career choice:

> *I love being a nurse and I love taking care of people; however, this pandemic has made me question my career. I've had a lot of time in isolation to think about it. As I long to hug my family but will only FaceTime them to keep them safe. I even had to watch my daughter's graduation online. Is it worth it to risk my life and my families for this career?*

Another nurse commented on feeling alone because of family not understanding what they are going through:

> *For those working third shift (like myself) how are you handling all this with your families??? Mine doesn't understand at all.. and I feel so utterly alone right now. They tell me that I "signed up for this job" so I'm not allowed to be saddened by it. I just don't know what to do, but I'm extremely depressed.*

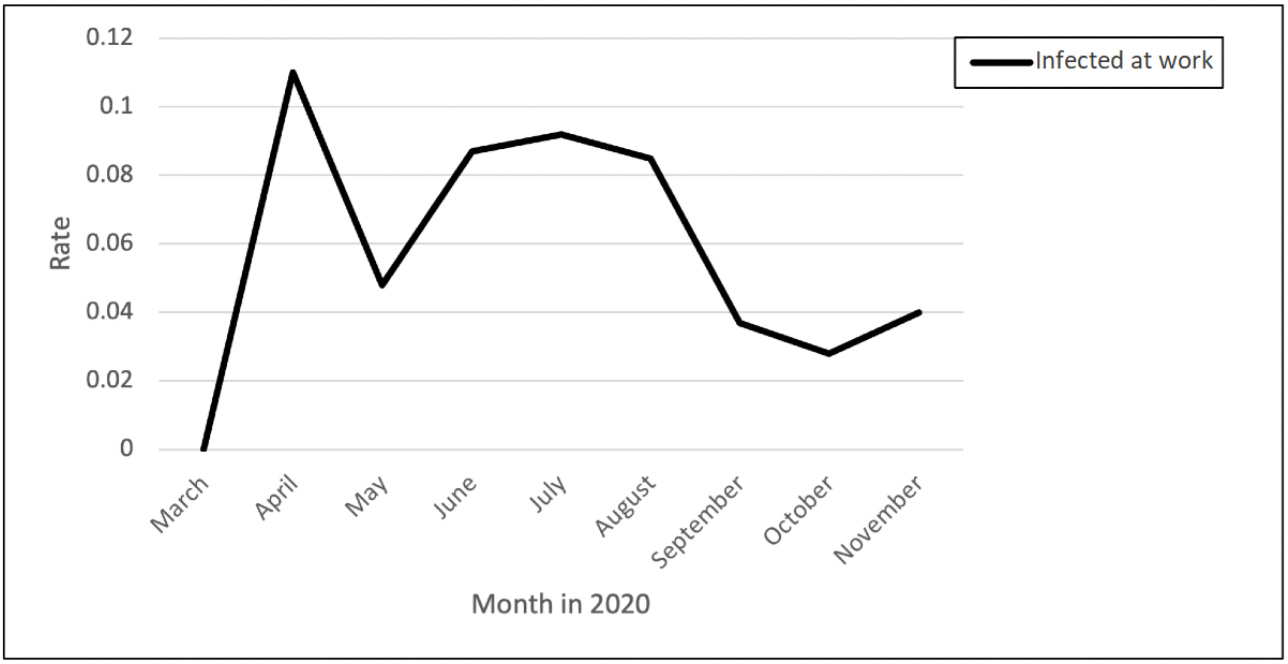
## Infected at Work

The posts and comments related to fear of getting infected peaked in April, followed by a decrease in the rate of posts and comments and another increase during June and July, after which they gradually decreased across the remaining months (Figure 7). An example post related to becoming infected at work was as follows:



XSL•FO
RenderX

**Figure 7.** Rates of posts and comments related to fear of getting infected at work over time.




> *We have had nearly 50 positive cases between staff and residents in our facility. No one is intentionally spreading it and we are all doing the best we can...Today was my absolute worst day ever in healthcare. I know everyone is under a lot of stress right now, but we all need to be a team. And when a teammate returns after an illness that landed them in the ER [emergency room] to go back at it and put themselves right back in the line of fire, have some compassion and show some respect!*

A nurse explained how she believes she got infected from the improper PPE they were given at work:

> *I believe I got infected with an ill-fitting KN95 back when we had to tape them to our faces. It makes me so damn angry that the US is the "richest" country in the world and yet PPE is still a problem nine months into this crud.*

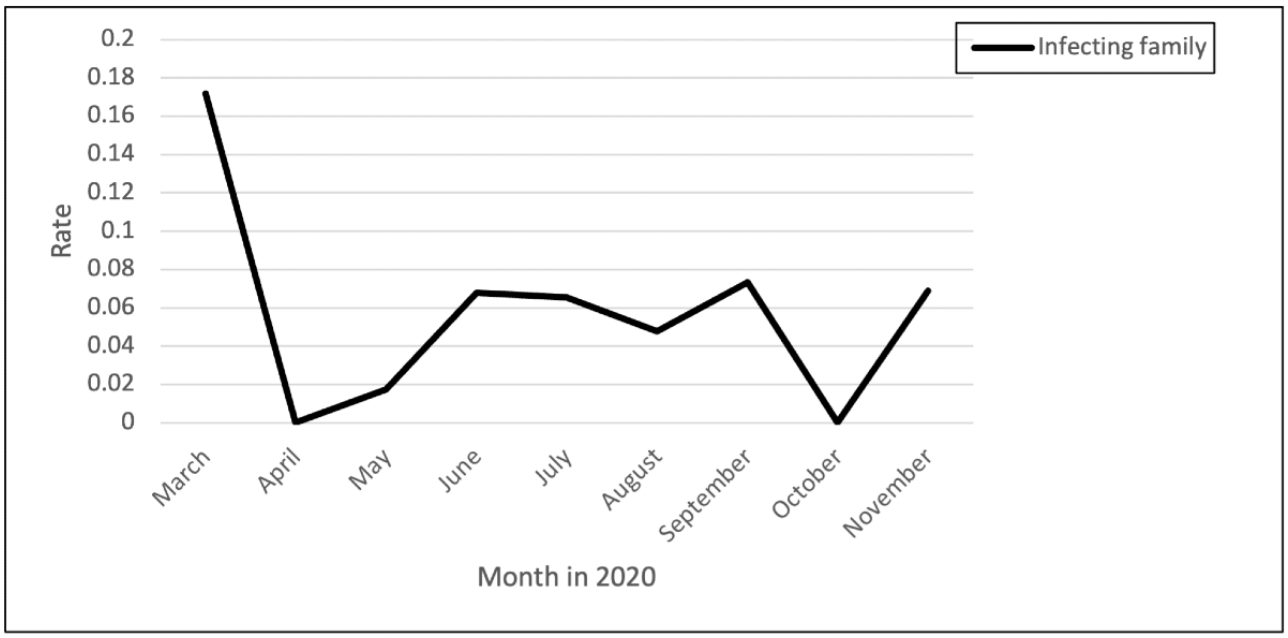
Another nurse discussed how mask wearing and social distancing are not followed in break rooms:

> *We have a very low number of positives. We have been doing masks and social distancing where required. At work we have to wear masks in the lab but when we hit the breakroom, masks come off and no social distancing.*

## Fear of Infecting Family

Nurses described the fear of infecting their family members who live in their home (Figure 8):

**Figure 8.** Rates of posts and comments related to fear of infecting family over time.






XSL•FO
RenderX

*...I think the ultimate challenge is protecting our families. I don't think the public totally gets the stress of how that burdens us.*

Another nurse described the fear of infecting their family and others as well as having to isolate from their children:

*...The increasing fear every day that I walk into the hospital that today is the day that someone who refuses to keep their mask on and coughs on me will give me the virus. I am afraid that I will be forced to self isolate and will have to explain to my small children why I can't give them hugs and kisses, or even come upstairs. I am afraid that I will unknowingly bring it home to my family, or my next patient that I come into contact with. I am afraid that if I do just one tiny thing wrong during the donning and doffing process, that I will be the reason someone gets sick...*

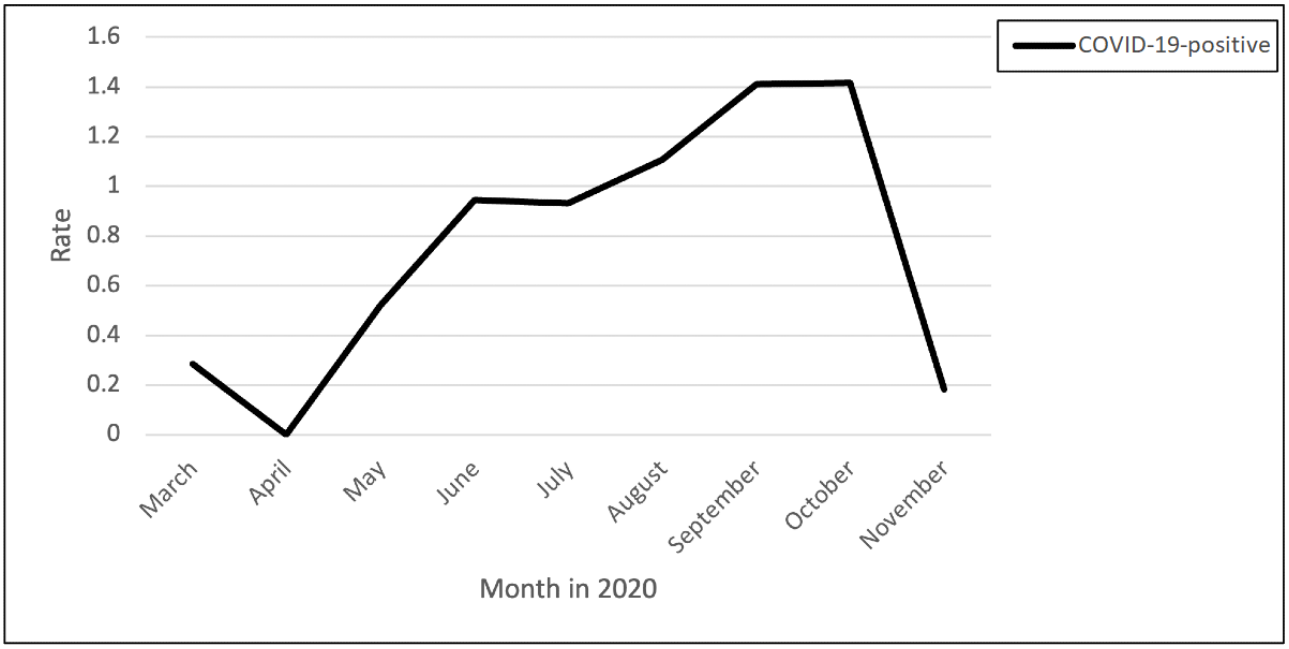
Another nurse explained how constantly changing protocols made her fearful of bringing COVID-19 home:

*...I definitely don't want to bring it home and infect my family. I just don't understand why the protocols seem to differ from day to day, and even hour by hour.*

## COVID-19–Positive Tests

The rate of posts regarding testing positive for COVID-19 peaked from April through October (Figure 9). Nurses infected with COVID-19 described symptoms they experienced. One nurse posted the following:

**Figure 9.** Rates of posts and comments related to testing positive for COVID-19 over time.




*Is there light at the end of the tunnel? On day 8 and day 6-7 I thought I was dying. Currently wondering if I should medicate for 100.5-degree temp or if it's better to let immune system fight it. Shortness of breath is better but cough is still there along with chills, body aches, diarrhea, stress incontinence from coughing so much. I was dizzy, had numbness and pins and needles in hands and feet and did nothing but sleep for 48 hours*

Symptoms from the virus lasted after recovery from the infection, compromising the ability to work:

*...19 days post symptom onset, went back to work...on Sunday to work three 12s in a row after being off for almost three weeks. I am EXHAUSTED, my brain is straight fog and I move so slow. My body kills and my feet are swollen. And I'm tachy with palpitations for 90% of the night unless I'm sitting for a long period of time which does not ever happen. I don't know how I will survive another night shift tonight. I can't breathe in my surgical mask, let alone my n95. My chest hurts from struggling to breathe through these shifts. I know it takes time to get completely back to normal but I am so frustrated and tired ...*

In addition, some of the nurses had long-lasting symptoms. One nurse posted the following in October:

*I had COVID in July and my sense of smell is not anywhere back to normal. When there is an odor, I smell the most rancid smell I could ever imagine. Anyone else experiencing this?! Will I ever go back to normal?!*

Exposure to the virus resulted in contracting COVID-19 and isolation from family:

*...have been isolated from my family for a week. I was diagnosed last Sunday. Breaks my heart that I can't see my children and I have to blow kisses to them from a screen. I tried my best to keep me and them safe. Praying for your health. This is no joke. I don't wish this upon anyone.*

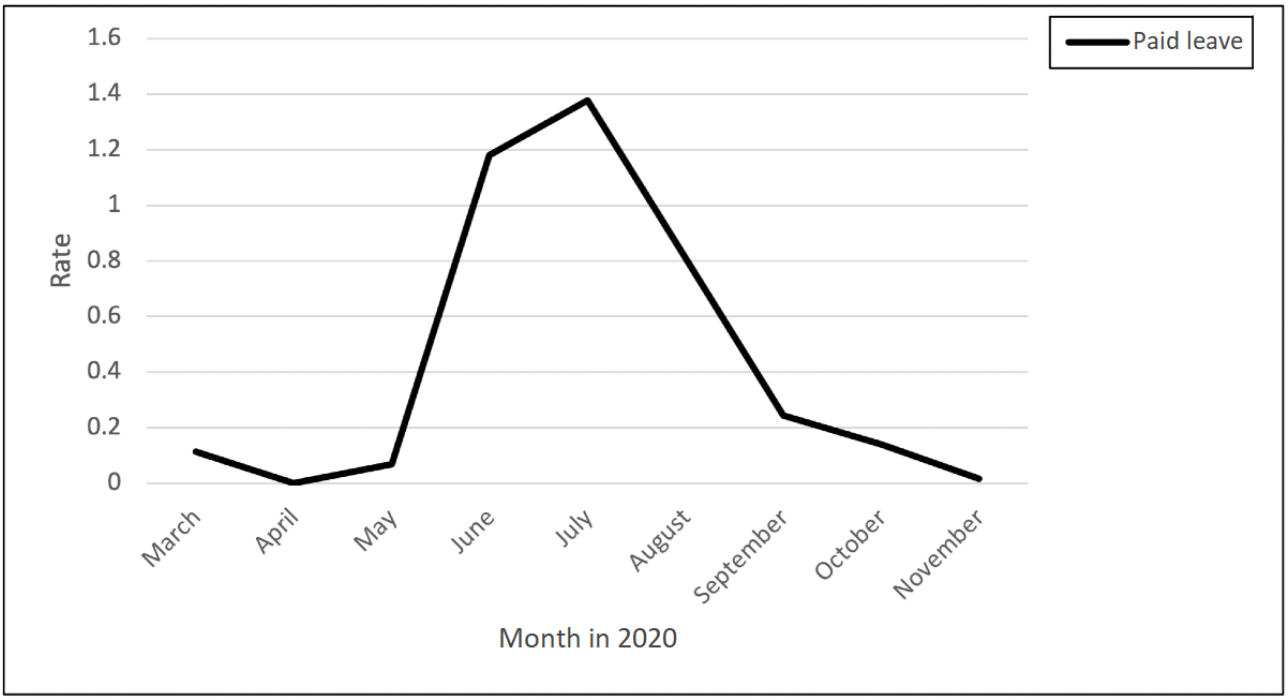


XSL•FO
RenderX

## Paid Leave

The rate of posts related to paid leave peaked from the beginning of May until mid-July and then declined through November (Figure 10). Nurses posted comments about their high-risk occupation that is not reflected in hazard pay. One of the nurses posted the following:

**Figure 10.** Rates of posts and comments related to paid leave due to COVID-19 over time.




> *Have been a nurse for 23 years and I agree with you. It's a shame that we are at such high risk, and the pay truly doesn't match the risk, not to mention the lack of pay when you finally test positive and have to stay home for 2-3 weeks (where I am now).*

Many nurses were struggling with unpaid leave that added an economic burden. Variability was seen among states regarding the policy of paid or nonpaid leave for nurses who tested positive for COVID-19. One nurse posted the following:

> *Just tested positive. Contracted it at work...I'm now home for 2 weeks, unpaid. Can someone help me understand how this is okay...My company doesn't have to compensate me despite contracting the virus while working. Tips? Ideas?*

## Detection of Positive Emotions Expressed by Nurses Over Time

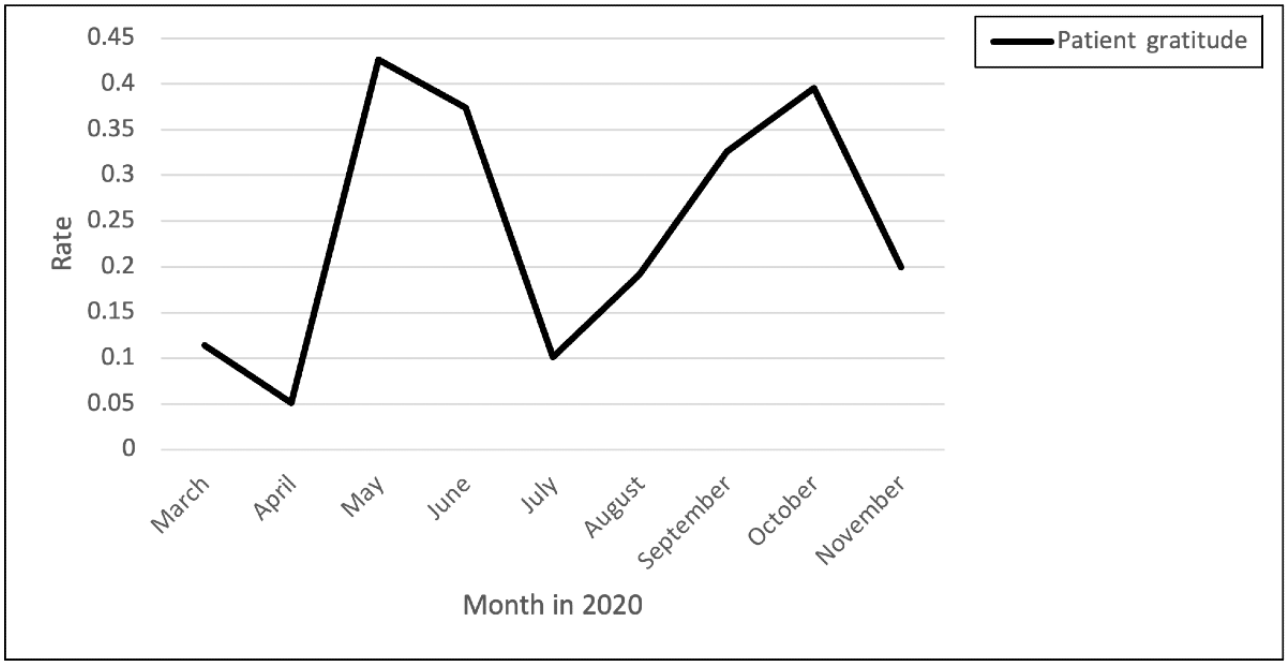
### *Patient Gratitude*

Posts and comments related to patient gratitude peaked from April through June and August through October (Figure 11). One of the nurses shared an example of appreciation expressions:



XSL•FO
RenderX

**Figure 11.** Rates of positive posts and comments related to patient gratitude over time.




*We had one of our nurses have her gas paid for by another patron. I have a friend who was given a gift card to Walmart when she was shopping. And I know someone who is giving the local hospital staff certificates for a free massage as they leave work...*

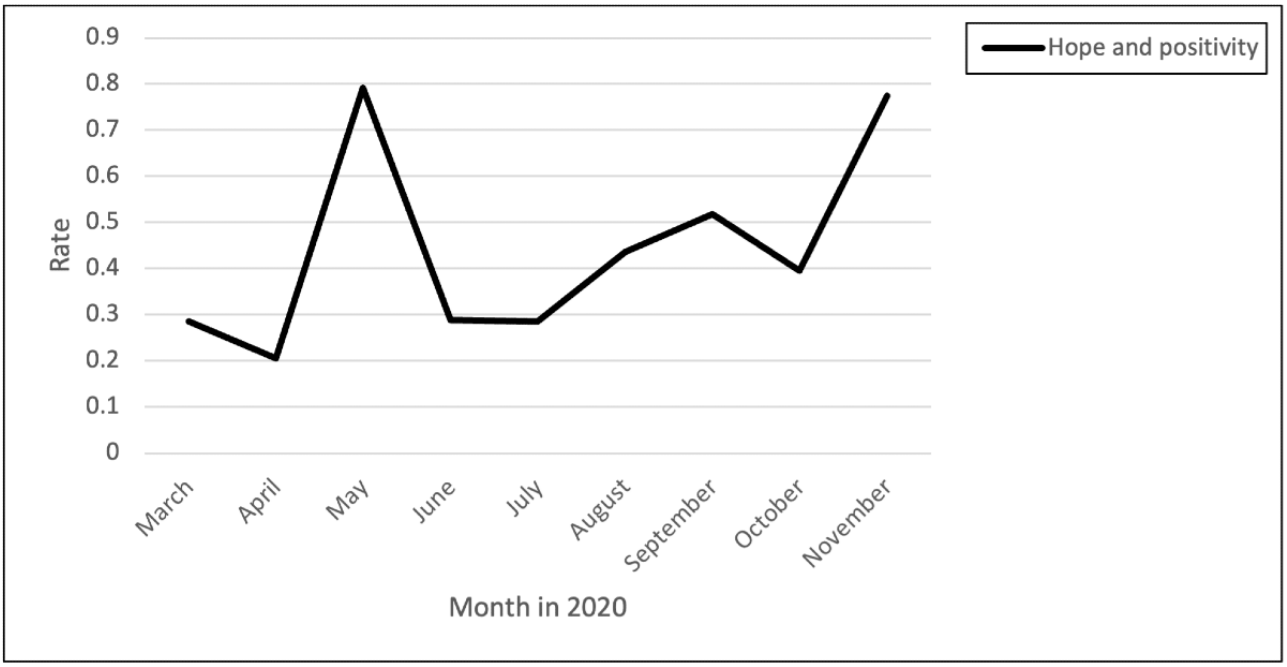
Another nurse commented the following:

*...this makes me so happy! I'm glad some people are appreciative.*

### *Hope*

Expressions of hope and positivity varied across the time periods, with the highest hope and positivity in May, followed by a smaller peak in September and then a continued increase from October through November (Figure 12). One comment represents the positivity and hope expressed as it relates to the strength gained from teamwork and not going through this pandemic alone:

**Figure 12.** Rates of "hope" comments over time.




*You are not alone. In your fear Frustration and Anger We see the tears wishing for better days You are not alone We see the strength Teamwork Desire to wrap your arms around a tired coworker Your loved ones The patient With no family near Pride in profession You are not alone We see you We are you.*



XSL•FO
RenderX

Many comments reflected hope that the peak effects of the pandemic would subside and that a return to some normalcy would be on the horizon:

> *Hopefully someday it will be less overwhelming.*
>
> *Our peak/wave is over here and hopefully never comes back.*

Other comments reflected positivity as nurses encouraged each other to keep hopeful:

> *You will come out of it soon!...Protect yourself at all cost. Most of all keep the faith. God bless you*

## Discussion

### Principal Findings

This social media study combined sentiment-detecting technology and major discussion themes to explore nurses' emotional expressions from the beginning of the pandemic through November 2020. The analyses were supported by direct quotes illuminating the experience of being a nurse during the COVID-19 pandemic. Our data methodologies follow the standard social media text analytic literature, which has been proven effective and trustworthy and has been applied in significant social media text analytics for nursing and COVID-19 trend studies in the past [36,38-45]. Apart from that, because of the inability of Facebook's API to extract the necessary group posts, we developed our own offline textual information extraction techniques, with appropriate deidentification to preserve users' privacy, as per IRB exemption conditions. Combining BERT-based sentiment analytic and Anserini-based information retrieval techniques facilitated the development of a full-fledged generalized social media analytics framework that can be used in any study domain that includes, but is not limited to, perspectives of students, media personalities, social workers, and minority groups with regard to adverse social events.

Sentiments described in the posts and comments reflected a variety of negative and positive emotions toward the pandemic experience. The negative sentiments expressed by nurses were anger, specifically as it relates to undefined policies, such as returning to work after being infected; anxiety because of being a frontline worker during the pandemic; and sadness caused by witnessing patients decline and die as well as being isolated. Nurses expressed mental and emotional exhaustion. The recent literature has expressed similar sentiments. In a cross-sectional descriptive correlative study on burnout of 2014 frontline nurses in Wuhan, China, 835 nurses reported high levels of emotional exhaustion, and 556 nurses experienced high level of depersonalization [8]. Sentiments related to anger and anxiety, specifically as they relate to undefined policies during nurses' service as frontline workers during the pandemic, were also expressed in recent studies. A commentary published by Nelson and Lee-Winn highlighted the anxiety nurses experienced as they dealt with very frequent changes in policies and protocols as the pandemic evolved [39]. Hu et al showed that 40% to 45% of frontline nurses experienced anxiety or depression, with 11% to 14% having moderate to severe anxiety or depression [8]. Results from a cross-sectional online survey of 1103 frontline emergency department nurses demonstrated that engaging in clinical services for patients with COVID-19 was significantly associated with a higher risk of depression (43.6%) [7].

Nurses shared factors that contributed to increased stress and anxiety. One example from this analysis was related to conspiracy theories and "fake news." Recent research has supported these findings [10]. Other contributing factors to the nurses' stress and anxiety were the shortage of PPE and noncompliance with rules on mask wearing. Nurses also reported development of skin lesions as a consequence of wearing PPE daily for long shifts. Similarly, Hu et al [8] and Shaukat et al [9] found that 1910 out of 2014 nurses had one or more skin lesions caused by PPE. Because of the shortage of PPE, the fear of becoming infected with COVID-19 as well as infecting their family members presented another factor in elevated anxiety and depression among the nurses. In addition, nurses expressed feelings of loneliness caused by isolation from their social life, family, and friends. Nelson and Lee-Win reported similar concerns. Similarly, a 2020 survey by the American Nurses Association of 10,997 nurses found that 28% felt depressed and 29% felt isolated and lonely [54].

Positive sentiments expressed by nurses were related to patient gratitude and hope as it related to teamwork and support of one another throughout the pandemic as well as hope for better days. Recent studies reported that nurses experienced positive emotions simultaneously with reporting negative emotions. There was a sense of responsibility and professional identity while they supported each other. The nurses also felt patients' gratitude [36]. These results agreed with our findings.

The emotions described also changed over time from the beginning of the pandemic until late November 2020. The findings resemble the psychosocial and emotional responses associated with the phases of disaster as described by DeWolfe [5]. Nurses spoke about their fears and anxiety, especially as they related to their sense of loss of ability to protect themselves and others, particularly their family members. These sentiments were noted throughout the time frame of posts and comments analyzed but were heightened in the early phase of the COVID-19 pandemic. This resembles Phase 1, or the predisaster phase, and the Phase 2 impact phase of a disaster as described by DeWolfe [5]. As the COVID-19 pandemic time frame continued (Multimedia Appendix 2), emotions experienced followed the Phase 2 impact and Phase 3 heroic phases of disaster but quickly moved to the Phase 5 disillusionment phase, as in this phase, the realization of limited assistance and noncompliance of the public led to emotions of stress and burnout, with many reactions such as exhaustion, frustration, anger, and depression being exhibited in the sentiments that were expressed. As vaccines were developed and the number of cases declined, positive emotions of gratitude and hope were displayed in the sentiments of posts and comments. This resembles the Phase 4 honeymoon phase. Specifically, optimistic comments were related to patient gratitude, teamwork, and support, as well as keeping the faith that all would return to normal.



XSL•FO
RenderX

## Limitations

Although this study was designed to be a unique representation of the perspective of nurses during the COVID-19 pandemic, there is a potential error in that some posts in this group could have been made by nonnurse individuals because of open-group posting allowance. Because of the specificity of comments and group rules that were monitored by administrators, the chances of this were low and presumably would not have affected the analysis. An additional limitation to this study is that the data originated from one specific open group on Facebook and might not represent all nurses' perspectives; however, in this one group, the membership included 106,000 nurses. In addition, findings from this study were in agreement with findings in the current published literature about this topic.

Although the authors based the analysis on the definitions described in Multimedia Appendix 3, they might not have captured all the emotions experienced by the nurses. We considered the pretrained BERT model, with four sentiment labels—joy, sadness, anger, and fear—which may slightly limit our analytic results. On the other hand, a single post that could be counted multiple times with regard to different sentiments carried the risk of introducing error into our analytic results. However, recent studies have found that considering only the above four sentiment labels and making multiple counts of the same posts for the different sentiments sustained the analytic results for COVID-19–related posts as per different machine learning techniques, which affirm the consistency of our results [55].

## Conclusions

The significance of this study is that it adds to the importance of documentation about a historical pandemic from the nurses' experience. The COVID-19 pandemic is a unique experience that the world was not prepared for and for which we were not preparing student nurses in the nursing curriculum. Themes and information gathered from this analysis will constitute evidence of what transpired in the United States in the time of the pandemic outbreak. It will provide a voice for the nurses who served on the front line. It will also serve as a basis for articulating lessons learned and a basis for ethical discussions of other topics in health care. In addition, it will be particularly useful to various government agencies, hospitals, organizations, and communities that wish to better understand the major concerns related to crises of public health and make policies to address them.

## Acknowledgments

This project has been supported by a UMass Lowell internal grant for COVID-19 impact analysis.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Topics and their associated phrases and methods used in this analysis.
[DOCX File , 17 KB - formative_v5i12e31358_app1.docx ]

Multimedia Appendix 2
COVID-19 timeline.
[DOCX File , 14 KB - formative_v5i12e31358_app2.docx ]

Multimedia Appendix 3
Number of posts by sentiments and themes over time.
[DOCX File , 19 KB - formative_v5i12e31358_app3.docx ]

## References

1. Institute of Medicine, Committee on Enhancing Environmental Health Content in Nursing Practice. In: Pope AM, Snyder MA, Mood LH, editors. Nursing, Health, and the Environment. Washington, DC: National Academies Press; 1995.
2. Haddad LM, Geiger RA. Nursing Ethical Considerations. Treasure Island, FL: StatPearls Publishing; 2021.
3. Morley G, Grady C, McCarthy J, Ulrich CM. Covid-19: Ethical challenges for nurses. Hastings Cent Rep 2020 May;50(3):35-39 [FREE Full text] [doi: 10.1002/hast.1110] [Medline: 32410225]
4. Cabarkapa S, Nadjidai SE, Murgier J, Ng CH. The psychological impact of COVID-19 and other viral epidemics on frontline healthcare workers and ways to address it: A rapid systematic review. Brain Behav Immun Health 2020 Oct;8:100144 [FREE Full text] [doi: 10.1016/j.bbih.2020.100144] [Medline: 32959031]
5. DeWolfe DJ. Training Manual for Mental Health and Human Service Workers in Major Disasters. 2nd edition. Rockville, MD: US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services; 2000. URL: https://files.eric.ed.gov/fulltext/ED459383.pdf [accessed 2021-11-21]
6. Aksoy YE, Koçak V. Psychological effects of nurses and midwives due to COVID-19 outbreak: The case of Turkey. Arch Psychiatr Nurs 2020 Oct;34(5):427-433 [FREE Full text] [doi: 10.1016/j.apnu.2020.07.011] [Medline: 33032769]



XSL•FO
RenderX

7. An Y, Yang Y, Wang A, Li Y, Zhang Q, Cheung T, et al. Prevalence of depression and its impact on quality of life among frontline nurses in emergency departments during the COVID-19 outbreak. J Affect Disord 2020 Nov 01;276:312-315 [FREE Full text] [doi: 10.1016/j.jad.2020.06.047] [Medline: 32871661]
8. Hu D, Kong Y, Li W, Han Q, Zhang X, Zhu LX, et al. Frontline nurses' burnout, anxiety, depression, and fear statuses and their associated factors during the COVID-19 outbreak in Wuhan, China: A large-scale cross-sectional study. EClinicalMedicine 2020 Jul;24:100424 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100424] [Medline: 32766539]
9. Shaukat N, Ali DM, Razzak J. Physical and mental health impacts of COVID-19 on healthcare workers: A scoping review. Int J Emerg Med 2020 Jul 20;13(1):40 [FREE Full text] [doi: 10.1186/s12245-020-00299-5] [Medline: 32689925]
10. Turale S, Meechamnan C, Kunaviktikul W. Challenging times: Ethics, nursing and the COVID-19 pandemic. Int Nurs Rev 2020 Jun;67(2):164-167 [FREE Full text] [doi: 10.1111/inr.12598] [Medline: 32578249]
11. Sun N, Wei L, Shi S, Jiao D, Song R, Ma L, et al. A qualitative study on the psychological experience of caregivers of COVID-19 patients. Am J Infect Control 2020 Jun;48(6):592-598 [FREE Full text] [doi: 10.1016/j.ajic.2020.03.018] [Medline: 32334904]
12. Liu Y, Zhai Z, Han Y, Liu F, Liu FP, Hu DY. Experiences of front-line nurses combating coronavirus disease-2019 in China: A qualitative analysis. Public Health Nurs 2020 Sep;37(5):757-763 [FREE Full text] [doi: 10.1111/phn.12768] [Medline: 32677072]
13. Abu-El-Noor NI, Abu-El-Noor MK. Ethical issues in caring for COVID-patients: A view from Gaza. Nurs Ethics 2020 Oct 14;27(8):1605-1606. [doi: 10.1177/0969733020956733]
14. He S, Ojo A, Beckman A, Gondi S, Gondi S, Betz M, et al. The story of #GetMePPE and GetUsPPE.org to mobilize health care response to COVID-19: Rapidly deploying digital tools for better health care. J Med Internet Res 2020 Jul 20;22(7):e20469 [FREE Full text] [doi: 10.2196/20469] [Medline: 32530813]
15. Lee S, Chen Y, Huang C. A text-driven rule-based system for emotion cause detection. In: Proceedings of the NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. 2010 Presented at: NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text; June 5, 2010; Los Angeles, CA p. 45-53 URL: https://aclanthology.org/W10-0206.pdf
16. Benali H, Wu C, Potdar V. Computational approaches for emotion detection in text. In: Proceedings of the 4th IEEE International Conference on Digital Ecosystems and Technologies. 2020 Presented at: 4th IEEE International Conference on Digital Ecosystems and Technologies; April 13-16, 2010; Dubai, United Arab Emirates p. 172-177 URL: https://www.researchgate.net/publication/224185404_Computational_approaches_for_emotion_detection_in_text [doi: 10.1109/dest.2010.5610650]
17. Hasan M, Rundensteiner E, Agu E. Automatic emotion detection in text streams by analyzing Twitter data. Int J Data Sci Anal 2018 Feb 9;7(1):35-51. [doi: 10.1007/s41060-018-0096-z]
18. Shivhare S, Khethawat S. Emotion detection from text. Comput Sci Inf Technol 2012:371-377 [FREE Full text] [doi: 10.5121/csit.2012.2237]
19. Kao E, Liu C, Yang T, Hsieh C, Soo V. Towards text-based emotion detection a survey and possible improvements. In: Proceedings from the International Conference on Information Management and Engineering. 2009 Presented at: International Conference on Information Management and Engineering; April 3-5, 2009; Kuala Lumpur, Malaysia p. 70-74. [doi: 10.1109/icime.2009.113]
20. Jain VK, Kumar S, Fernandes SL. Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. J Comput Sci 2017 Jul;21:316-326. [doi: 10.1016/j.jocs.2017.01.010]
21. Desmet B, Hoste V. Emotion detection in suicide notes. Expert Syst Appl 2013 Nov;40(16):6351-6358. [doi: 10.1016/j.eswa.2013.05.050]
22. Joshi A, Tripathi A, Soni R, Bhattacharyya P, Carman M. EmoGram: An open-source time sequence-based emotion tracker and its innovative applications. In: Proceedings of the Workshops at the 30th AAAI Conference on Artificial Intelligence. 2016 Presented at: Workshops at the 30th AAAI Conference on Artificial Intelligence; February 12-13, 2016; Phoenix, AZ p. 512-516.
23. Mohammad SM, Kiritchenko S. Using hashtags to capture fine emotion categories from tweets. Comput Intell 2014 Jan 10;31(2):301-326. [doi: 10.1111/coin.12024]
24. Hasan M, Rundensteiner E, Agu E. Detecting emotions in twitter messages. In: Proceedings of the ASE Big Data/Social Computing/Cybersecurity Conference. 2014 Presented at: Proceedings of the ASE Big Data/Social Computing/Cybersecurity Conference; May 27-31, 2014; Stanford, CA p. 27-31 URL: https://web.cs.wpi.edu/~emmanuel/publications/PDFs/C30.pdf
25. Agrawal A, An A. Unsupervised emotion detection from text using semantic and syntactic relations. In: Proceedings of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. 2012 Presented at: IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology; December 4-7, 2012; Macau, China p. 346-353. [doi: 10.1109/wi-iat.2012.170]
26. Yasmina D, Hajar M, Hassan AM. Using YouTube comments for text-based emotion recognition. Procedia Comput Sci 2016;83:292-299. [doi: 10.1016/j.procs.2016.04.128]




XSL•FO
RenderX

27. Tiwari S, Vijaya RM, Phonsa G, Deepu D. A novel approach for detecting emotion in text. Indian J Sci Technol 2016;9(29):1-5. [doi: 10.17485/ijst/2016/v9i29/88211]
28. Ghazi D, Inkpen D, Szpakowicz S. Prior and contextual emotion of words in sentential context. Comput Speech Lang 2014 Jan;28(1):76-92. [doi: 10.1016/j.csl.2013.04.009]
29. Grover S, Verma A. Design for emotion detection of Punjabi text using hybrid approach. In: Proceedings of the International Conference on Inventive Computation Technologies. 2016 Presented at: International Conference on Inventive Computation Technologies; August 26-27, 2016; Coimbatore, India p. 26-27. [doi: 10.1109/inventive.2016.7824823]
30. Ekman P. An argument for basic emotions. Cogn Emot 2008 Jan 07;6(3-4):169-200. [doi: 10.1080/02699939208411068]
31. Shaver P, Schwartz J, Kirson D, O'Connor C. Emotion knowledge: Further exploration of a prototype approach. J Pers Soc Psychol 1987;52(6):1061-1086. [doi: 10.1037/0022-3514.52.6.1061]
32. Oatley K, Johnson-Laird PN. Towards a cognitive theory of emotions. Cogn Emot 1987 Mar;1(1):29-50. [doi: 10.1080/02699938708408362]
33. Plutchik R. Emotion: A Psychoevolutionary Synthesis. New York, NY: Harper & Row; 1980.
34. Russell JA. A circumplex model of affect. J Pers Soc Psychol 1980;39(6):1161-1178. [doi: 10.1037/h0077714]
35. Ortony A, Clore G, Collins A. The Cognitive Structure of Emotions. Cambridge, UK: Cambridge University Press; 1988.
36. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.18653/v1/N19-1423]
37. Go A, Bhayani R, Huang L. Twitter Sentiment Classification Using Distant Supervision. CS224N Project Report. Stanford, CA: Stanford University; 2009. URL: https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf [accessed 2021-11-21]
38. Yang P, Fang H, Lin J. Anserini: Reproducible ranking baselines using Lucene. J Data Inf Qual 2018 Nov 03;10(4):1-20. [doi: 10.1145/3239571]
39. Nelson SM, Lee-Winn AE. The mental turmoil of hospital nurses in the COVID-19 pandemic. Psychol Trauma 2020 Aug;12(S1):S126-S127. [doi: 10.1037/tra0000810] [Medline: 32584109]
40. Al Maskari Z, Al Blushi A, Khamis F, Al Tai A, Al Salmi I, Al Harthi H, et al. Characteristics of healthcare workers infected with COVID-19: A cross-sectional observational study. Int J Infect Dis 2021 Jan;102:32-36 [FREE Full text] [doi: 10.1016/j.ijid.2020.10.009] [Medline: 33039607]
41. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2000;1:993-1022 [FREE Full text]
42. Slavik CE, Buttle C, Sturrock SL, Darlington JC, Yiannakoulias N. Examining tweet content and engagement of Canadian public health agencies and decision makers during COVID-19: Mixed methods analysis. J Med Internet Res 2021 Mar 11;23(3):e24883 [FREE Full text] [doi: 10.2196/24883] [Medline: 33651705]
43. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
44. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
45. Lyu JC, Luli GK. Understanding the public discussion about the Centers for Disease Control and Prevention during the COVID-19 pandemic using Twitter data: Text mining analysis study. J Med Internet Res 2021 Feb 09;23(2):e25108 [FREE Full text] [doi: 10.2196/25108] [Medline: 33497351]
46. Devlin J, Chang MW. Open sourcing BERT: State-of-the-art pre-training for natural language processing. Google AI Blog. 2018 Nov 02. URL: https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html [accessed 2021-09-01]
47. Singh M, Jakhar AK, Pandey S. Sentiment analysis on the impact of coronavirus in social life using the BERT model. Soc Netw Anal Min 2021;11(1):33 [FREE Full text] [doi: 10.1007/s13278-021-00737-z] [Medline: 33758630]
48. Kaliyar RK, Goswami A, Narang P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimed Tools Appl 2021 Jan 07:1-24 [FREE Full text] [doi: 10.1007/s11042-020-10183-2] [Medline: 33432264]
49. Dai X, Karimi S, Hachey B, Paris C. Cost-effective selection of pretraining data: A case study of pretraining BERT on social media. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2020 Presented at: Conference on Empirical Methods in Natural Language Processing; November 16-20, 2020; Virtual p. 1675-1681 URL: https://aclanthology.org/2020.findings-emnlp.151.pdf [doi: 10.18653/v1/2020.findings-emnlp.151]
50. Azzopardi L, Moshfeghi Y, Halvey M, Alkhawaldeh RS, Balog K, Di Buccio E, et al. Lucene4IR: Developing information retrieval evaluation resources using Lucene. Workshop report. In: Proceedings of the ACM SIGIR Forum. 2017 Presented at: ACM SIGIR Forum; September 8-9, 2016; Glasgow, UK p. 58-75 URL: https://sigir.org/wp-content/uploads/2017/01/p058.pdf [doi: 10.1145/3053408.3053421]
51. Grand A, Muir R, Ferenczi J, Lin J. MAXSCORE to block-max WAND: The story of how Lucene significantly improved query evaluation performance. In: Proceedings of the European Conference on Information Retrieval. 2020 Presented at:




XSL•FO
RenderX

European Conference on Information Retrieval; April 14-17, 2020; Lisbon, Portugal p. 20-27. [doi: 10.1007/978-3-030-45442-5_3]
52. Kamphuis C, de Vries AP, Boytsov L, Lin J. Which BM25 do you mean? A large-scale reproducibility study of scoring variants. In: Proceedings of the European Conference on Information Retrieval. 2020 Presented at: European Conference on Information Retrieval; April 14-17, 2020; Lisbon, Portugal p. 28-34. [doi: 10.1007/978-3-030-45442-5_4]
53. Transformers. Hugging Face. 2020. URL: https://huggingface.co/transformers/ [accessed 2021-06-08]
54. American Nurses Foundation. Pulse on the Nation's Nurses COVID-19 Survey Series: Mental Health and Wellness Survey 1. ANA Enterprise COVID-19 Resource Center. 2020 Aug. URL: https://www.nursingworld.org/practice-policy/work-environment/health-safety/disaster-preparedness/coronavirus/what-you-need-to-know/mental-health-and-wellbeing-survey/ [accessed 2021-06-08]
55. Rustam F, Khalid M, Aslam W, Rupapara V, Mehmood A, Choi GS. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. PLoS One 2021;16(2):e0245909 [FREE Full text] [doi: 10.1371/journal.pone.0245909] [Medline: 33630869]

## Abbreviations

**API:** application programming interface
**BERT:** Bidirectional Encoder Representations from Transformers
**ER:** emergency room
**ICU:** intensive care unit
**IRB:** Institutional Review Board
**NLP:** natural language processing
**PPE:** personal protective equipment

*Edited by G Eysenbach; submitted 22.06.21; peer-reviewed by A Chen, R Sun, C Luo; comments to author 04.08.21; revised version received 07.09.21; accepted 26.09.21; published 10.12.21.*

*Please cite as:*
*Koren A, Alam MAU, Koneru S, DeVito A, Abdallah L, Liu B*
*Nursing Perspectives on the Impacts of COVID-19: Social Media Content Analysis*
*JMIR Form Res 2021;5(12):e31358*
*URL: https://formative.jmir.org/2021/12/e31358*
*doi:10.2196/31358*
*PMID:34623957*

©Ainat Koren, Mohammad Arif Ul Alam, Sravani Koneru, Alexa DeVito, Lisa Abdallah, Benyuan Liu. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Patients' Perspectives on Qualitative Olfactory Dysfunction: Thematic Analysis of Social Media Posts

Jane K Parker[1], BSc, PhD; Christine E Kelly[1,2], BA, MSc; Barry C Smith[3], MA, PhD; Aidan F Kirkwood[1], BSc, MRes; Claire Hopkins[4], BMBCh, MA, DM (Oxon); Simon Gane[5], MBChB, MPhil, DOHNS

[1]Department of Food and Nutritional Sciences, School of Chemistry, Food and Pharmacy, University of Reading, Reading, United Kingdom
[2]AbScent, Andover, United Kingdom
[3]Centre for the Study of the Senses, Institute of Philosophy, School of Advanced Study, University of London, London, United Kingdom
[4]Ear, Nose and Throat Department, Guy's and St Thomas' Hospitals, London, United Kingdom
[5]Royal National Ear, Nose and Throat and Eastman Dental Hospitals, London, United Kingdom

**Corresponding Author:**
Simon Gane, MBChB, MPhil, DOHNS
Royal National Ear, Nose and Throat and Eastman Dental Hospitals
47-49 Huntley Street
London, WC1E 6DG
United Kingdom
Phone: 44 7789655222
Email: simongane@nhs.net

## Abstract

**Background:** The impact of qualitative olfactory disorders is underestimated. Parosmia, the distorted perception of familiar odors, and phantosmia, the experience of odors in the absence of a stimulus, can arise following postinfectious anosmia, and the incidences of both have increased substantially since the outbreak of COVID-19.

**Objective:** The aims of this study are to explore the symptoms and sequalae of postinfectious olfactory dysfunction syndrome using unstructured and unsolicited threads from social media, and to articulate the perspectives and concerns of patients affected by these debilitating olfactory disorders.

**Methods:** A thematic analysis and content analysis of posts in the AbScent Parosmia and Phantosmia Support group on Facebook was conducted between June and December 2020.

**Results:** In this paper, we identify a novel symptom, olfactory perseveration, which is a triggered, identifiable, and usually unpleasant olfactory percept that persists in the absence of an ongoing stimulus. We also observe fluctuations in the intensity and duration of symptoms of parosmia, phantosmia, and olfactory perseveration. In addition, we identify a group of the most common items (coffee, meat, onion, and toothpaste) that trigger distortions; however, people have difficulty describing these distortions, using words associated with disgust and revulsion. The emotional aspect of living with qualitative olfactory dysfunction was evident and highlighted the detrimental impact on mental health.

**Conclusions:** Qualitative and unsolicited data acquired from social media has provided useful insights into the patient experience of parosmia and phantosmia, which can inform rehabilitation strategies and ongoing research into understanding the molecular triggers associated with parosmic distortions and research into patient benefit.

***(JMIR Form Res 2021;5(12):e29086)*** doi:10.2196/29086

**KEYWORDS**

olfactory dysfunction; parosmia; phantosmia; olfactory perseveration; trigger foods; mental health; COVID-19; patients' perspective; thematic analysis; social media; perspective; smell; nose; symptom; concern; support



XSL•FO
RenderX

## Introduction

### Background: Olfactory Dysfunction Before and Since COVID-19

Until recently, olfactory dysfunction was a little-recognized and underestimated disorder, distressing to those affected and with few effective treatments available. Prior to the COVID-19 pandemic, olfactory dysfunction was believed to affect approximately 19% of the general population, rising to 40% in those aged over 60 years [1]. The etiologies most commonly reported were sinonasal disease, upper respiratory tract infection, and traumatic brain injury [2].

Since the outbreak of COVID-19, cases of olfactory dysfunction have increased. The most recent estimate for COVID-19–related loss of smell and taste is 50% to 65% of all cases [3,4]. Given that there have been >200 million cases of COVID-19 globally (as of August 13, 2021) [5], approximately 100 million people will have been affected by smell loss. Although many will recover within weeks [6], it is estimated that approximately 10% will have long-term olfactory problems, many of whom will subsequently develop a qualitative olfactory dysfunction [7].

### Qualitative Olfactory Dysfunction

The term *qualitative olfactory dysfunction* covers both parosmia (qualitative distortion in the presence of an odor) and phantosmia (odor experience in the absence of odor).

Parosmia is the triggered (requiring an external stimulus), subjective perception of a qualitatively altered odor identity with a negative hedonic component (almost always unpleasant), and it usually subsides within seconds of the stimulus. It often develops in the early stages of recovery from smell loss, particularly after postinfectious and posttraumatic anosmia (loss of any sense of smell), with onset weeks after the initial insult. Those severely affected by parosmia find many familiar foods intolerable and start to reject food, leading to weight change, anxiety, and in severe cases, clinical depression [8-10]. Parosmia has been reported in 34% of all patients presenting with olfactory disorders (n=392), and the most common etiology is upper respiratory infection, with 56% of those with anosmia progressing to parosmia [11].

Phantosmia often occurs alongside parosmia [12] and is similarly a perception of an unpleasant subjective odor; however, it is not triggered by obvious external odorants. Patients experience many of the same objectionable odors that are perceived by those with parosmia [2], but these sensations can persist for days.

### Role of Social Media

Awareness of smell disorders has undoubtedly increased since the start of the COVID-19 pandemic, after it became recognized worldwide as one of the key symptoms [13]. However, relatively few publications are dedicated to understanding either the pathophysiology of the disease and its impact on the patient. The charity AbScent (Registration No. 1183468 in England and Wales) has provided support for those with such disorders, launching several support groups on Facebook, the first in 2015 [14]. A COVID-19 Smell and Taste Loss group was started in March 2020 [15], and once it became apparent that post–COVID-19 anosmics were also developing parosmia, a group dedicated to those experiencing parosmia and phantosmia was started in June 2020 [16]. This group accumulated >7000 followers by February 2021 and is an important source of information for researchers, providing valuable insight into the nature and progression of the disease. The fact that people are turning to social media for support is evidence that patients are in need of more information to help them understand both the disease and the efficacy of various treatments, as well as finding social interaction and moral support in coping with an often debilitating condition.

### Aims of This Study

The aims of the paper are to qualitatively explore the symptoms and sequalae of postinfectious olfactory dysfunction syndrome, using unstructured and unsolicited threads from social media, and to articulate the perspectives and concerns of patients affected by these debilitating olfactory disorders.

## Methods

### Approach

The use of social media is rarely discussed in the published literature; however, Alanin et al [17] argue strongly that this approach, when driven by the patients' own perspectives, provides invaluable, unsolicited, and spontaneous data that would not otherwise be retrieved from more structured surveys and questionnaires. It paints a rich picture of the patient journey, avoiding bias arising from the structure of the survey and the generation of artefacts where patients are eager to please. Other biases may be introduced with this method; for example, the sample is not randomized, and no conclusions about incidence or prevalence can be made due to self-selection.

### Source of Data

The findings reported here are taken from posts made in the AbScent Parosmia and Phantosmia Support closed group on Facebook between June 12 and December 14, 2020 [16]. Demographic data are not available, as those joining the group do not provide this information. Conversation within this group is lively, and responses to polls and questions can generate over 300 comments in 24 hours. Although much of the research into the changes in smell and taste during the pandemic has focused on patients who have had confirmed cases of COVID-19, either through positive tests or clinical diagnosis, the information presented here is about recent self-reported changes that have not been confirmed as being related to COVID-19 infection.

This study has been approved by the University of Reading School of Chemistry, Food and Pharmacy Research Ethics Committee (study number 39.2020), and express permission was sought through messaging channels from each person quoted.

### Thematic Analysis

Thematic analysis was conducted by following the approach used by Alanin et al [17]. An inductive approach was used to allow the data to determine the themes, and a flexible reflexive



XSL•FO
RenderX

thematic analysis was used to code the threads. The Facebook group moderator (CK) monitored the discussion daily, noting recurring themes, and a second researcher (JP) independently reviewed the posts to identify themes. The themes were reviewed, and after discussion, 7 major themes were identified by consensus and subthemes emerged. Quotes that were deemed to illustrate the major ideas in each theme or subtheme, and for which permission for inclusion had been freely obtained, were selected for presentation (verbatim) in Table S1, Multimedia Appendix 1.

### Content Analysis

To identify the foods most often associated with parosmia, the data from one thread of conversation were collated manually and counted twice. The conversation was prompted by the moderator, who posed the question: "Can you all add here your worst foods for parosmia?" (137 comments). Data were collected during the first 3 months of the study using a simple frequency table to record each food as it was mentioned (Table S2, Multimedia Appendix 1). In terms of text analysis, this is a very small data set, and manual coding was deemed appropriate, although the repeatability and tolerance associated with this method should be borne in mind.

## Results

### Theme 1. "How It Makes Me Feel"

An extensive and concerning theme demonstrated the detrimental effect of parosmia and phantosmia on emotional well-being. Members of the Facebook group reported being sad and miserable (comments 1-5 in Table S1, Multimedia Appendix 1). Some were quite scared and terrified by the prospect of having parosmia (6-9) and found reassurance when they found others in a similar position (10-13). Patients felt alone and isolated (14-15) and were relieved to find the support of the online community. Some were frustrated with a lack of understanding from others (16) and responses from their physicians, who seemed uninterested (17-18) and lacked treatment options. Many reported a decline in their mental health (19-23), and one member was surprised that loss of smell could have such an impact on their mental health (24). Lack of the ability to smell body odor was a worry, as was the inability, from a safety perspective, to smell smoke or gas. Some adjusted to the "new normal" (25), and many were looking for hope (26-28). Although early in the process, some posted about their recovery (29-33), which was usually only partial. Few mentioned complete recovery, but by this stage, these users might have left the group without commenting further.

### Theme 2. Fluctuations

One emerging theme is the fluctuating nature of the symptoms during the recovery process. Often, users experienced partial recovery (hyposmia), and many reported a complete return of a normal sense of smell before the onset of parosmia (34). The duration of symptoms varied, as some had occasional "whiffs" of normality occurring during long stretches of parosmia (35). For others, the changes in intensity and the magnitude of the fluctuations were quite extreme (36-38) and persistent (38), and the severity of parosmia and phantosmia (92-100) could vary daily. Both hormones (38) and tiredness or stress (39) were associated with more severe fluctuations.

### Theme 3. Items That Trigger Parosmic Experiences

Much of the online discussion concerned problematic foods, drinks, or household items, as patients started to recognize and record their own experiences and explore whether the same items elicited the same response in others. Some listed up to 15 to 20 different items, including food, drinks, bleach, cigarettes, and personal care items (shower gel, shampoo, and hand sanitizers). Many mentioned coffee to be one of the worst triggers (40-41). Fried, toasted, or roasted foods were common triggers of distortion (42-44), as were chocolate (45-46) and onions and garlic (47-49); however, these were not universal. Long lists of other foods were provided that included carbohydrates, fruit, vegetables, herbs, spices (50), and even water (51); thus, it was almost impossible to create a list of foods that never triggered parosmic sensations.

Personal care items were frequently mentioned (42-55); many users attributed their parosmia to different brands or ingredients and started exploring alternative products. Toothpaste and mint were frequent offenders (56-59), and there were recommendations to switch to alternative flavored toothpastes (60).

### Theme 4. Defining the Character of Parosmic Distortions

One defining feature of parosmia is the distortion of (familiar) smells, which members had difficulty describing because they could not relate the smell to a previous experience (61-65). Although many descriptions were used to describe "that parosmia smell," they were often prefixed with "it's like" in an attempt to describe the associated disgust rather than the smell identity. To many, it was an entirely novel smell. A group of frequently used words seemed to be based on a burnt, chemical, or dirty connotation: burnt cigarettes, burnt rubber, sewage, earthy, dirty, rancid, death and decay, or unpleasant (66-70). However, there was evidence of at least one other type of parosmia smell, with many reporting their triggers as sickly sweet and rotten (71-73). These two different concepts were often attributed to different foods (74-75) but had been used together to describe one parosmia smell (76) or in a progression in which the burnt character appeared first but changed into the sweeter smell after several months (77).

The feeling of disgust and revulsion associated with the distortions (78-79) was clear and could induce vomiting in a handful of cases (80-82). Disgust was not always mentioned explicitly, but many implied their disgust by their choice of words, which were associated with disgust (garbage, sewage, decay, feces).

### Theme 5. The Smell of Feces

This was a recurrent theme in the Facebook group. If perceived at all, the smell of feces was usually more pleasant than expected (83-85) and often took on the same character as food, such as onion and garlic, whether this was a distorted (86) or normal (87) onion and garlic smell. The smell of feces was often described as distorted coffee (88-89) or as a sweet smell (90-91).



XSL•FO
RenderX

The striking corollary of this finding is that the hedonic value of these odors was reversed: odors that typically elicit disgust were less objectionable than before, but odors that usually had a positive hedonic value were perceived as disgusting, and the confusion between food smells and bodily waste products could cause additional distress.

### Theme 6. Phantosmia

Many group members confused parosmia and phantosmia (92), not knowing whether what they perceived was real (93). In general, phantosmia was discussed less frequently, even though it caused much anxiety, as the perceived odor could last for days (94), weeks (95), or even months. Similar to parosmia, the descriptions of the smell were cigarette, chemical, burnt and rotting (96-97), or sweet and sickly (99), and the phantosmia was subject to fluctuations (96, 98). An unusual and novel finding was that in some cases, members described it as a triggered reaction (100-101); this was echoed strongly in the AbScent COVID-19 Smell and Taste Loss group, where there was a thread dedicated to phantosmia.

### Theme 7. Tips and Tricks for Survival

Evidently, the members of the group provided significant support for each other (102-103), creating a positive environment with few negative comments. Some posted practical tips to mitigate the impact (104-106), and many were keen to pass on their own experience and provide lists of "safe" foods that cause minimal distortions. In general, it is unsurprising that appetite was lost (107), as bland foods were frequently recommended ("the plainer the better," 108-109)) and many turned to fresh fruit and vegetables (110), plain carbohydrates (111-112), and dairy products (112-115) as "safe" options. Cooked and roasted foods that tend to be a major source of protein (meat, nuts) were not considered "safe," and it is concerning that some experienced difficulty finding a palatable source of protein (116-117). Whereas some found acceptable alternatives (Quorn, turkey mince, and protein shakes), others reported diets lacking in anything nutritious (118-119) or abstained from food completely (120), putting themselves at risk of malnutrition.

### Content Analysis

More than 75 different items were mentioned as triggers of a parosmic sensation in a conversation prompted by the moderator. By using our selection criteria (avoiding multi-ingredient foods, focusing on simple ingredients, and combining where necessary), a list of 50 items that trigger distortions, and the frequency with which they were cited, is shown in Table S2. The four major food triggers, each cited >40 times, were coffee, meat, onions, and toothpaste/mint, followed by garlic and eggs, which were cited >20 times. Personal care products such as shower gel, deodorant, soap, shampoo, and hand sanitizer were often triggers, and these were gathered into a single category and were mentioned >30 times.

## Discussion

### Principal Results

A total of 7 themes emerged from our analysis, as outlined below:

- Theme 1 confirms that the negative emotional and mental health impacts of smell dysfunction, already described in the literature [8,9,18,19], are prominent in those experiencing parosmia and phantosmia.
- In Theme 2, we uncover the fluctuating nature and duration of the symptoms of parosmia and phantosmia during the recovery process.
- In Theme 3, we identify a group of the most common items (coffee, meat, onion, and toothpaste) that trigger distortions, and this was supported by a quantitative content analysis of the threads.
- In Theme 4, we demonstrate how difficult it is for those with parosmia to describe their perception of distorted items, and that most describe their hedonic reaction to the distortions rather than describing the characteristics of the smell.
- In Theme 5, we collate people's responses to fecal smells. We observe a hedonic reversal in that fecal smells are often reported to be more pleasant (or less unpleasant) than before olfactory loss or change.
- In Theme 6, the focus is on phantosmia, or distortions that are untriggered; however, we observe what we believe is a novel symptom, which we term "olfactory perseveration"—a triggered, identifiable, and usually unpleasant olfactory percept that persists, sometimes for days, in the absence of an ongoing stimulus.
- Theme 7 identifies tips and tricks for survival as provided by the members of the group.

In the discussion that follows, we have synthesized the thematic findings to discuss the emotional and clinical manifestations of qualitative disorders, common triggers, and coping strategies.

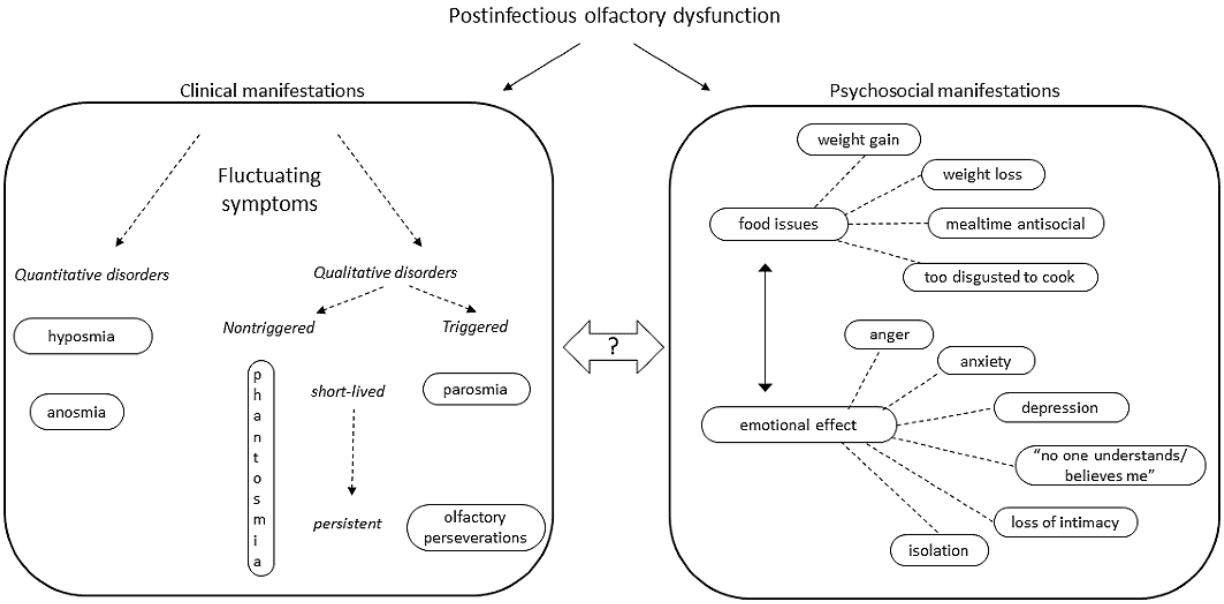
### Emotional Manifestations

The emotional and mental health impacts of smell disorders was a major topic of discussion in the group. Figure 1 summarizes the psychosocial manifestations, incorporating observations on food issues from Burges Watson et al [19] showing the interdependence of food issues and emotional effects. Smell loss or change is acutely felt by those suddenly confronted with the way their experience of the world is altered in terms of a lack of pleasure in eating and the absence of reassuring smells of familiar people and places. Many feel socially isolated and go on to experience long-lasting depression.



XSL•FO
RenderX

**Figure 1.** Clinical and psychosocial manifestations of olfactory dysfunction, incorporating conclusions from Burges Watson et al [19].




## Clinical Manifestations

Accounts of clinical manifestations are summarized in Figure 1, which shows an undefined yet complex and fluctuating relationship between different olfactory disorders (Theme 2). Contrary to what we have learned anecdotally from non–COVID-19 incidences of parosmia, posts in this recent group indicate that parosmia can be preceded by the almost complete return of a normal sense of smell. However, it was demonstrated recently that there is a mismatch between subjective self-reported olfactory function and objective olfactory tests: more than half of 112 patients reporting a normal sense of smell were found to have a quantitative loss of sense of smell, of whom 6 were anosmic [20]. During further progression of the disease, experiences of the symptoms of olfactory disorders vary in intensity and duration (prolonged periods or short whiffs), giving episodes of different qualitative olfactory disorders and intermittent returns to anosmia or even to what is perceived as normal olfactory function. These fluctuations have no obvious basis or sequence, can occur in series or in parallel, and vary from case to case. Given the close association between olfactory function and mental well-being [8], it is conceivable that fluctuations may be influenced by emotional status.

What was surprising in Theme 6 was the elucidation of what we believe to be a novel symptom: "smell lock" or olfactory perseveration. A search of the literature finds only one mention of this phrase, in a paper on olfactory hallucinations in Alzheimer disease [21], where it was not further described. We propose a definition of this symptom as a triggered, identifiable, usually unpleasant olfactory percept that persists in the absence of an ongoing stimulus. We suspect that it is often confused with phantosmia and is relatively common in postinfectious smell loss syndrome. It is distinct from phantosmia in that it is triggered by a recognizable stimulus and remains identifiable as the smell of that trigger while persisting long after the original stimulus is removed from the surrounding environment. Describing novel symptoms such as this both increases our understanding of patient experience and guides further research efforts.

## Common Triggers

The wide range of trigger foods and personal care products identified in Theme 3 is consistent with a survey of 725 patients conducted by Keller and Malaspina [2]. This study also reported coffee as an "efficient" trigger, and it was often mentioned together with chocolate. Meat, coffee, and cocoa represent some of the most complex aroma profiles, as the exposure of the combination of sugars, amino acids, and fats to high temperatures (coffee roasting, roasting of cocoa nibs, roasting meat) often produces a rich mix of aroma compounds. Because these and other cooked foods listed (fried foods, peanut butter, bacon, and toast) are frequent triggers that are rarely reported as "safe foods," it is reasonable to suggest that they may have aroma compounds in common that trigger parosmia. Onion and garlic are mentioned far less frequently in the Keller and Malaspina study but are major triggers of parosmic distortions in this study. Both contain a variety of potent sulfur compounds, which we propose may be responsible for initiating the experience of parosmia. Many unheated foods are also triggers (bell pepper, citrus, apple, cucumber, and banana); therefore, parosmia is not simply associated with cooked foods. Mint and toothpaste also seem to be powerful triggers, although it is not yet clear whether this may also be due to stimulation of the trigeminal nerve. Personal care products are objectionable to many, but these are very brand dependent and subject to the ingredients, particularly the choice of essential oils and flavorings used by different manufacturers.

In Theme 5, we find that the smell of feces takes on the character of other distorted foods, suggesting that the strongly



XSL•FO
RenderX

objectionable odors typically associated with feces are not being perceived. There is a curious switching in hedonic valence between odors that are usually objectionable and those that are usually highly desirable: it seems that "fair is foul, and foul is fair" in parosmia.

### Coping Strategies

Advice on food is crucial from nutritional, hedonic, and emotional perspectives (Theme 7). Avoidance of the top triggers (coffee, meat, eggs, onion, garlic, and toothpaste) and most roasted or baked foods makes sense; however, this may lead to nutritional deficiencies, especially in those who rely on meat and eggs as their main sources of protein. Furthermore, avoidance of triggers may hinder the adaptation process that occurs, albeit very slowly. This is one area where further research would be beneficial. It is far better for those afflicted to develop coping strategies for unpalatable foods. This could involve minimizing thermal load; boiling or steaming rather than roasting or frying; and minimizing flavor release by consumption of foods that are chilled or at room temperature. What is clear is that every person is different, and people need to experiment to find a varied diet with the right balance of nutrition and reward.

## Conclusions

Qualitative methods are sometimes perceived as being less rigorous than quantitative methods; however, we have shown how valuable such methods can be in both understanding patients' lived experiences and constraining hypotheses for quantitative investigation. By collation and interpretation of themes derived from social media, we have articulated the concerns and experiences of patients and gained a better understanding of qualitative olfactory dysfunction and the stimuli that trigger the responses of disgust, learning over time how the disease progresses and how it affects eating behavior, social interactions, and mental health. Forums such as the AbScent Facebook group provide much-needed moral support, where members immediately feel connected, listened to, and relieved to know that they are not alone. This itself can lift spirits, representing an important first step in coming to terms with what is, for many, a relatively long-term condition. The value of patient-centered research should not be underestimated. This qualitative research, based on thematic analysis of unsolicited social media posts, provides the framework for ongoing investigation and therapeutic intervention in smell disorders, shedding further light on the impact of these problems and guiding the direction of further research.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 45 KB - formative_v5i12e29086_app1.docx ]

## References


1. Brämerson A, Johansson L, Ek L, Nordin S, Bende M. Prevalence of olfactory dysfunction: the skövde population-based study. Laryngoscope 2004 Apr;114(4):733-737. [doi: 10.1097/00005537-200404000-00026] [Medline: 15064632]
2. Keller A, Malaspina D. Hidden consequences of olfactory dysfunction: a patient report series. BMC Ear Nose Throat Disord 2013 Jul 23;13(1):8 [FREE Full text] [doi: 10.1186/1472-6815-13-8] [Medline: 23875929]
3. Borsetto D, Hopkins C, Philips V, Obholzer R, Tirelli G, Polesel J, et al. Self-reported alteration of sense of smell or taste in patients with COVID-19: a systematic review and meta-analysis on 3563 patients. Rhinology 2020 Oct 01;58(5):430-436. [doi: 10.4193/Rhin20.185] [Medline: 32626853]
4. Parma V, Ohla K, Veldhuizen MG, Niv MY, Kelly CE, Bakke AJ, GCCR Group Author, et al. More than smell - COVID-19 is associated with severe impairment of smell, taste, and chemesthesis. Chem Senses 2020 Oct 09;45(7):609-622 [FREE Full text] [doi: 10.1093/chemse/bjaa041] [Medline: 32564071]
5. Coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-08-13]
6. Boscolo-Rizzo P, Polesel J, Spinato G, Menegaldo A, Fabbris C, Calvanese L, et al. Predominance of an altered sense of smell or taste among long-lasting symptoms in patients with mildly symptomatic COVID-19. Rhinology 2020 Oct 01;58(5):524-525. [doi: 10.4193/Rhin20.263] [Medline: 32683438]
7. Hopkins C, Surda P, Vaira LA, Lechien JR, Safarian M, Saussez S, et al. Six month follow-up of self-reported loss of smell during the COVID-19 pandemic. Rhinology 2021 Feb 01;59(1):26-31. [doi: 10.4193/Rhin20.544] [Medline: 33320115]
8. Croy I, Nordin S, Hummel T. Olfactory disorders and quality of life-an updated review. Chem Senses 2014 Mar;39(3):185-194. [doi: 10.1093/chemse/bjt072] [Medline: 24429163]
9. Kohli P, Soler ZM, Nguyen SA, Muus JS, Schlosser RJ. The association between olfaction and depression: a systematic review. Chem Senses 2016 Jul;41(6):479-486 [FREE Full text] [doi: 10.1093/chemse/bjw061] [Medline: 27170667]
10. Leopold D. Distortion of olfactory perception: diagnosis and treatment. Chem Senses 2002 Sep;27(7):611-615. [doi: 10.1093/chemse/27.7.611] [Medline: 12200340]
11. Reden J, Maroldt H, Fritz A, Zahnert T, Hummel T. A study on the prognostic significance of qualitative olfactory dysfunction. Eur Arch Otorhinolaryngol 2007 Feb;264(2):139-144. [doi: 10.1007/s00405-006-0157-0] [Medline: 17006637]




XSL•FO
RenderX

12. Hummel T, Whitcroft KL, Andrews P, Altundag A, Cinghi C, Costanzo RM, et al. Position paper on olfactory dysfunction. Rhinol Suppl 2017 Mar;54(26):1-30. [doi: 10.4193/Rhino16.248] [Medline: 29528615]
13. Gane SB, Kelly C, Hopkins C. Isolated sudden onset anosmia in COVID-19 infection. A novel syndrome? Rhinology 2020 Jun 01;58(3):299-301. [doi: 10.4193/Rhin20.114] [Medline: 32240279]
14. AbScent | Anosmia Support. Facebook. URL: https://www.facebook.com/groups/AbScent [accessed 2021-10-30]
15. Covid-19 Smell and Taste Loss. Facebook. URL: https://www.facebook.com/groups/AbScentCovid19 [accessed 2021-10-30]
16. AbScent Parosmia and Phantosmia Support. Facebook. URL: https://www.facebook.com/groups/AbScentParosmia [accessed 2021-10-30]
17. Alanin MC, Laidlaw T, Society TS, Hopkins C. The burden of non-steroidal anti-inflammatory exacerbated respiratory disease from the patient's perspective - a qualitative analysis of posts from the Samter's Society. Rhinology 2020 Aug 01;58(4):333-340. [doi: 10.4193/Rhin19.430] [Medline: 32219224]
18. Soudry Y, Lemogne C, Malinvaud D, Consoli S, Bonfils P. Olfactory system and emotion: common substrates. Eur Ann Otorhinolaryngol Head Neck Dis 2011 Jan;128(1):18-23 [FREE Full text] [doi: 10.1016/j.anorl.2010.09.007] [Medline: 21227767]
19. Burges Watson D, Campbell M, Hopkins C, Smith B, Kelly C, Deary V. Altered smell and taste: anosmia, parosmia and the impact of long Covid-19. PLoS One 2021;16(9):e0256998 [FREE Full text] [doi: 10.1371/journal.pone.0256998] [Medline: 34559820]
20. Boscolo-Rizzo P, Menegaldo A, Fabbris C, Spinato G, Borsetto D, Vaira L, et al. High prevalence of long-term psychophysical olfactory dysfunction in patients with COVID-19. medRxiv ;46:bjab006 Preprint posted online on January 08, 2021. [FREE Full text] [doi: 10.1101/2021.01.07.21249406]
21. El Haj M, Larøi F. Olfactory hallucinations in Alzheimer's disease. Acta Neuropsychiatr 2021 Feb;33(1):37-42. [doi: 10.1017/neu.2020.33] [Medline: 32981566]

*Edited by G Eysenbach; submitted 25.03.21; peer-reviewed by K Koly, HM Tzeng; comments to author 31.07.21; revised version received 26.08.21; accepted 18.09.21; published 14.12.21.*

*Please cite as:*
*Parker JK, Kelly CE, Smith BC, Kirkwood AF, Hopkins C, Gane S*
*Patients' Perspectives on Qualitative Olfactory Dysfunction: Thematic Analysis of Social Media Posts*
*JMIR Form Res 2021;5(12):e29086*
*URL: https://formative.jmir.org/2021/12/e29086*
*doi:10.2196/29086*
*PMID:34904953*

©Jane K Parker, Christine E Kelly, Barry C Smith, Aidan F Kirkwood, Claire Hopkins, Simon Gane. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Predictors to Use Mobile Apps for Monitoring COVID-19 Symptoms and Contact Tracing: Survey Among Dutch Citizens

Stephanie Jansen-Kosterink[1,2*], PhD; Marian Hurmuz[1,2*], MSc; Marjolein den Ouden[3*], PhD; Lex van Velsen[1,2*], PhD

[1]eHealth Department, Roessingh Research and Development, Enschede, Netherlands

[2]Biomedical Signals and Systems Group, University of Twente, Enschede, Netherlands

[3]Technology, Health & Care, Saxion University of Applied Sciences, Enschede, Netherlands

[*]all authors contributed equally

**Corresponding Author:**
Stephanie Jansen-Kosterink, PhD
eHealth Department
Roessingh Research and Development
Roessinghsbleekweg 33b
Enschede, 7522 AH
Netherlands
Phone: 31 88 0875 777
Email: s.jansen@rrd.nl

## Abstract

**Background:** eHealth apps have been recognized as a valuable tool to reduce COVID-19's effective reproduction number. The factors that determine the acceptance of COVID-19 apps remain unknown. The exception here is privacy.

**Objective:** The aim of this article was to identify antecedents of acceptance of (1) a mobile app for COVID-19 symptom recognition and monitoring and (2) a mobile app for contact tracing, both by means of an online survey among Dutch citizens.

**Methods:** Next to the demographics, the online survey contained questions focusing on perceived health, fear of COVID-19, and intention to use. We used snowball sampling via posts on social media and personal connections. To identify antecedents of the model for acceptance of the 2 mobile apps, we conducted multiple linear regression analyses.

**Results:** In total, 238 Dutch adults completed the survey; 59.2% (n=141) of the responders were female and the average age was 45.6 years (SD 17.4 years). For the symptom app, the final model included the predictors age, attitude toward technology, and fear of COVID-19. The model had an r2 of 0.141. The final model for the tracing app included the same predictors and had an r2 of 0.156. The main reason to use both mobile apps was to control the spread of the COVID-19 virus. Concerns about privacy was mentioned as the main reason to not use the mobile apps.

**Conclusions:** Age, attitude toward technology, and fear of COVID-19 are important predictors of the acceptance of COVID-19 mobile apps for symptom recognition and monitoring and for contact tracing. These predictors should be taken into account during the development and implementation of these mobile apps to secure acceptance.

***(JMIR Form Res 2021;5(12):e28416)*** doi:10.2196/28416

**KEYWORDS**

COVID-19; eHealth; mHealth; contact tracing; symptom management; intention to use

## Introduction

It is spring 2020 and the COVID-19 pandemic has the world in its grip. Infection with COVID-19 can lead to a simple cold or no symptoms at all, while it can also rapidly develop into a life-threatening disease, especially for patients with existing cardiovascular problems, obesity, or diabetes [1]. To hamper the spread of COVID-19 and to manage the intensive care unit capacity, many countries have applied a lockdown strategy for their citizens [2]. In order to control the spread of COVID-19 after a lockdown, and to minimize the effective reproduction number of the disease, several measures can be applied, of which social distancing, combined with aggressive case finding and isolation, seems to be the most effective [3].

eHealth apps have been recognized as a valuable tool for supporting symptom recognition and monitoring [4], for contact tracing [5], and ultimately, for reducing COVID-19's effective reproduction number by means of timely intervention. In short,



XSL•FO
RenderX

a contact tracing app would record a citizen's contacts with other people via Bluetooth technology and, in the case of a COVID-19 infection, will warn the persons that the index patient recently had contact with so that they can apply self-isolation and be attentive for any COVID-19 symptoms. However, for such apps to be effective, high uptake among the population is necessary. For the case of a tracing app, it has been estimated that 56% of a country's population should use the app to suppress the epidemic [6]. It is therefore crucial that the design of these apps and the implementation strategies that accompany them take the factors that affect acceptance into account.

The factors that determine acceptance of COVID-19 apps are largely unknown [7]. The exception here is privacy. Since the initial plan of governments to implement these technologies, a fierce public debate erupted on whether or not large-scale tracing of contacts for this goal is an unacceptable breach of privacy or not. While the issue of privacy has been recognized as an important antecedent of acceptance of mobile health apps [8], the unique and disturbing situation that the COVID-19 pandemic places us in makes it difficult to apply existing models and frameworks for eHealth acceptance. In May 2020 the Dutch government wanted to develop and implement 2 mobile apps to prevent the spread of the COVID-19 virus and support Dutch municipal health services. The aim of this article was to identify antecedents of acceptance of (1) a mobile app for COVID-19 symptom recognition and monitoring, and (2) a mobile app for contact tracing, both by means of an online survey among Dutch citizens.

# Methods

## Overview

To identify antecedents of acceptance of a mobile app for COVID-19 symptoms recognition and monitoring (hereafter: symptom app), and a mobile app for contact tracing (hereafter: tracing app), an online survey was developed, tested, and distributed among Dutch citizens. This study did not require formal ethical approval (as ruled by CMO Oost Nederland, file number: 2020-6628). At the beginning of the survey, participants were asked for consent to use their data for research purposes.

## Survey

### Design

The online survey (Multimedia Appendix 1) consisted of 4 parts. The first part included questions on demographics, the second part contained questions related to perceived health, the third part consisted of questions related to the fear of a COVID-19 infection, and the final part included questions to assess the intention to use the 2 suggested mobile apps. In April 2020, the Dutch government announced plans to develop and implement 2 mobile apps for preventing the spread of the COVID-19 virus. However, the exact design of these apps remained unknown at this time. Therefore, we introduced both mobile apps in the survey via a short description of their general aim. We pretested the survey with 14 Dutch citizens to improve legibility.

### Demographics

We assessed gender, age, smartphone use, educational level (student, primary school, secondary school, high school, bachelor's degree/university/PhD), work status (unemployed and searching for work, not able to work due to illness, volunteer work, part-time work, full-time work, retired, student), income level (below-average wages, average wages, above-average wages), and living status (living alone, living together, other). We assessed the participants' attitude toward technology using the Personal Innovativeness in the Domain of Information Technology scale by Agarwal and Prasad [9], consisting of 4 statements and accompanied by a 5-point Likert scale (ranging from 1 [strongly disagree] to 5 [strongly agree]). Finally, we also asked whether participants were (once) infected with COVID-19. The answer options for this question were: Yes, In doubt, or No.

### Perceived Health

To assess perceived health, we asked participants to complete 3 questions. These questions were used previously to assess perceived health among Dutch citizens [10]. These questions/statements were (1) How would you describe your health?; (2) How concerned are you about your health?; and (3) I am ill more often than other people of the same age and sex. These were accompanied by a 5-point Likert scale ranging from 1 (bad, not concerned, and totally disagree, respectively) to 5 (excellent, very concerned, and totally agree, respectively).

### Fear of COVID-19

The participants' fear of a COVID-19 infection was assessed by means of 4 questions related to this topic:

- Have you been concerned about the outbreak of the COVID-19 virus in recent weeks?: 5-point Likert scale, ranging from 1 (not at all concerned) to 5 (extremely concerned)
- How often did you think of the outbreak of the COVID-19 virus in recent weeks?: 5-point Likert scale, ranging from 1 (never) to 5 (always)
- How afraid were you of the outbreak of the COVID-19 virus in recent weeks?: 5-point Likert scale, ranging from 1 (not afraid at all) to 5 (very afraid)
- How afraid are you of getting sick from the COVID-19 virus?: 5-point Likert scale, ranging from 1 (not afraid at all) to 5 (very afraid)

### Intention to Use

Finally, participants were asked to rate their intention to use the 2 mobile apps: (1) a symptom app and (2) a tracing app. The statements for the construct intention to use were based on van Velsen et al [11]. All 3 questions were accompanied by a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Next to these closed questions, respondents were also asked what the main reasons were to "use" and "not to use" the mobile apps.

## Survey Distribution

Distribution of the survey (via QualtricsXM) started on April 15, 2020. Participants were eligible if they were 18 years of age or older. We used a snowball sampling via posts on social media



XSL•FO
RenderX

(LinkedIn, Twitter, and Facebook) and personal connections. Next to this, we recruited participants via a Dutch panel of older adults that indicated they were interested in participating in research on the topic of eHealth. The survey was closed on April 30, 2020. Due to the method of recruitment, a response rate could not be calculated.

### Analyses

Data were analyzed using SPSS (version 19; IBM). Descriptive statistics were performed for all outcomes. Cronbach α values were calculated to assess internal consistency for attitude toward technology, perceived health, fear of COVID-19, and intention to use. Next, survey scores were interpreted for these factors as being negative (score 1 or 2), neutral (score of 3), or positive (score 4 or 5). Via a paired (2-tailed) *t* test, the difference in intention to use score between both mobile apps was tested. To identify antecedents of acceptance of (1) a symptom app and (2) a tracing app, we conducted multiple linear regression analyses (backward model analyses). The intention to use each app was used as the dependent variable. The independent variables were selected based on Pearson correlation coefficients. Demographic characteristics and factors that (borderline) significant correlated (Pearson correlation cut-off level $P \leq .10$) with the dependent variable "intention to use" were included in the multiple linear regression analyses. For the paired *t* test and regression analyses, the level of significance was set at $P < .05$. For the final models the $r^2$ was calculated, which indicates the percentage of the variance in the dependent variable that the independent variables explain collectively. To support the quantitative results, the responses on the 2 open questions were sorted and counted by the first author and discussed with the second author, taking an inductive approach. Disagreements were discussed until unanimous agreement was reached.

## Results

### Composition of Survey Participants

In total, 238 Dutch citizens completed the survey. Fifteen responders only completed the intention to use survey of a tracing app as this app was presented first and these responders stopped with the survey after these questions; 59.2% (141/238) of the responders were female and the average age was 45.6 years (SD 17.4 years). Only 2.1% (5/238) of responders did not own a smartphone and 74.8% (178/238) claimed that they carried their smartphone with them for most of the day. The average age of our sample is higher than the average age of the Dutch population. There was also an overrepresentation of female participants and participants with a high education level [12]. Compared with the statistics of 2018, in the current sample, there is an overrepresentation of participants owning a smartphone and compared with the statistics of 2020 there is an underrepresentation of participants who are unemployed [12]. The internal consistency of the Attitude Toward Technology scale was good (Cronbach α=.85). Most responders (176/238, 73.9%) had a moderate attitude toward technology. Only 3/238 responders (1.3%) claimed to be infected with COVID-19. All demographic characteristics are presented in Table 1.



XSL•FO
RenderX

**Table 1.** Responders' demographics (n=238).

| Demographics | Values |
|---|---|
| **Gender, n (%)** | |
| Male | 97 (40.8) |
| Female | 141 (59.2) |
| Age (years), mean (SD) | 45.6 (17.4) |
| **Smartphone, n (%)** | |
| Yes | 233 (97.9) |
| No | 5 (2.1) |
| **Carry smartphone with you?, n (%)** | |
| Always | 178 (74.8) |
| Sometimes | 55 (23.1) |
| Never | 5 (2.1) |
| **Education level, n (%)** | |
| Student | 16 (6.7) |
| Primary school | 2 (0.8) |
| Secondary school | 14 (5.9) |
| High school | 57 (23.9) |
| Bachelor's degree/university/PhD | 149 (62.6) |
| **Work status, n (%)** | |
| Unemployed and searching for work | 3 (1.3) |
| Not able to work due to illness | 8 (3.4) |
| Volunteer work | 2 (0.8) |
| Part-time work | 75 (31.5) |
| Full-time work | 81 (34.0) |
| Retired | 43 (18.1) |
| Student | 25 (10.5) |
| **Income level, n (%)** | |
| Below-average wages | 76 (31.9) |
| Average wages | 93 (39.1) |
| Above-average wages | 69 (29.0) |
| **Living status, n (%)** | |
| Living alone | 34 (14.3) |
| Living together | 191 (80.3) |
| Other | 13 (5.5) |
| **Attitude toward technology,[a] mean (SD)** | 3.2 (0.78) |
| Low (1-2) | 3 (1.3) |
| Moderate (3) | 176 (73.9) |
| High (4-5) | 59 (24.8) |
| **COVID-19 infection, n (%)** | |
| Yes | 3 (1.3) |
| In doubt | 44 (18.5) |
| No | 191 (80.3) |



XSL•FO
RenderX

[a]Measured with a scale from 1 (low) to 5 (high).

## Fear of a COVID-19 Infection

The internal consistency of the 4 items in this scale was acceptable to good (Cronbach α=.78). The mean score on this topic was 3.3 (SD 0.68). The majority of the responder's opinion on this topic was neutral (192/238, 80.7%) and 16% (38/238) of the responders were afraid for a COVID-19 infection. Only a few responders (8/238, 3.4%) were not afraid (Table 2).

**Table 2.** Descriptive statistics and internal consistency of scales.

| Scale | Number of items | Cronbach α | Mean (SD) | Positive, n (%) | Neutral, n (%) | Negative, n (%) |
|---|---|---|---|---|---|---|
| Fear of COVID-19 (n=238) | 4 | .78 | 3.3 (0.68) | 38 (16.0) | 192 (80.7) | 8 (3.4) |
| Perceived health (n=238) | 3 | .69 | 3.8 (0.68) | 139 (58.4) | 97 (40.8) | 2 (0.8) |
| Intention to use the symptom app (n=223) | 3 | .96 | 3.38 (1.07) | 101 (45.3) | 101 (45.3) | 21 (9.4) |
| Intention to use the tracing app (n=238) | 3 | .96 | 3.27 (1.14) | 98 (41.2) | 108 (45.4) | 32 (13.4) |

## Perceived Health

For the 3 items to assess the perceived health of the responders the internal consistence was acceptable (Cronbach α=.69). The mean score on this scale was 3.8 (SD 0.68). Most respondents were positive about their health (139/238, 58.4%).

## Intention to Use

The intention to use was assessed for the symptom app and the tracing app. For both scales, internal consistency was excellent (Cronbach α for the symptom app=.96 and Cronbach α for the tracing app=.96). For both apps, the majority's intention to use was neutral (Table 2). However, an additional paired *t* test indicated that there was a significant difference in the scores on intention to use for the symptom app (mean 3.38 [SD 1.07]; n=223) and the tracing app (mean 3.27 [SD 1.13]; n=223; $t_{222}$=–2.598 and *P*=.01), indicating that the responders were more willing to use a mobile app for COVID-19 symptom recognition and monitoring compared with a mobile app for contact tracing.

## Correlations

The intention to use the symptom app was related to income level (r=0.132, *P*=.05), attitude toward technology (r=0.220, *P*<.001), and fear of COVID-19 (r=–0.291, *P*<.001). The intention to use the tracing app was related to age (r=0.135, *P*=.04), attitude toward technology (r=0.223, *P*<.001), and fear of COVID-19 (r=–0.303, *P*<.001). Based on these outcomes, the independent variables within the linear regression analysis were age, income level, attitude toward technology, fear of COVID-19, and perceived health. Table 3 provides an overview of the correlations between all demographics and factors, and the intention to use.



XSL•FO
RenderX

**Table 3.** Outcome Pearson correlation.

| Variables | Intention to use the symptom app (n=223) | Intention to use the tracing app (n=238) |
|---|---|---|
| Gender | r=–0.056<br>*P*=.41 | r=–0.147<br>*P*=.23 |
| Age | r=0.126<br>*P*=.06 | r=0.135[a]<br>*P*=.03 |
| Education level | r=0.21<br>*P*=.76 | r=0.018<br>*P*=.79 |
| Work status | r=0.072<br>*P*=.28 | r=0.033<br>*P*=.62 |
| Income level | r=–0.132[a]<br>*P*=.05 | r=0.124<br>*P*=.06 |
| Living status | r=0.083<br>*P*=.22 | r=0.060<br>*P*=.35 |
| Attitude toward technology | r=0.220[a]<br>*P*<.001 | r=0.223[a]<br>*P*<.001 |
| Fear of COVID-19 | r=–0.291[a]<br>*P*<.001 | r=–0.303[a]<br>*P*<.001 |
| Perceived health | r=–0.088<br>*P*=.19 | r=–0.119<br>*P*=.07 |

[a]Correlation is significant at the .05 level (2-tailed).

## Linear Regression

A multiple linear regression analysis was conducted to predict the intention to use a symptom app based on age, income level, attitude toward technology, fear of COVID-19, and perceived health. The final model included the predictors attitude toward technology, fear of COVID-19, and age ($F_{3,3}$=12.012; *P*<.001).

The model has an $r^2$ of 0.141. It contains 3 factors that affect the intention to use, but only 2 of them are significant predictors:

- Fear of COVID-19: β=–.272, $t_3$=4.305, *P*<.001
- Attitude toward technology: β=.222, $t_3$=3.532, *P*=.001
- Age: β=.107, $t_3$=1.691, *P*=.09 (not significant)

Another multiple linear regression analysis was conducted to predict the intention to use a tracing app based on age, income level, attitude toward technology, fear of COVID-19, and perceived health. The final model included the predictors attitude toward technology, fear of COVID-19, and age ($F_{3,3}$=14.333; *P*<.001). The model has an $r^2$ of 0.155. Intention to use is predicted by:

- Fear of COVID-19: β=.286, $t_3$=4.742, *P*<.001
- Attitude toward technology: β=.230, $t_3$=3.815, *P*<.001
- Age: β=.128, $t_3$=2.104, *P*<.05

## Main Reason to Use the Mobile Apps

An overview of all reasons the responders brought forth for using both mobile apps is presented in Tables 4 and 5. The main reason (33/116, 28.4) for responders to use the symptom app was to control the spread of the COVID-19 virus. In addition, respondents were willing to use this mobile app to monitor own complaints (22/116, 19.0%) and to gain more insight into the spread and symptoms of the COVID-19 virus (19/116, 16.4%).

The main reason to use a tracing app was also to control the spread of the COVID-19 virus (45/147, 30.6%). Next to this, respondents were willing to use this mobile app to gain more insight into the spread and symptoms of the COVID-19 virus (34/147, 23.1%) and for one's own health (19/147, 12.9%).



XSL•FO
RenderX

**Table 4.** Overview of the main reasons to use the symptom app (n=116).

| Reasons | Value, n (%) |
|---|---|
| To control the spread of the COVID-19 virus in general | 33 (28.4) |
| To monitor own complaints | 22 (19.0) |
| More insight into the spread and symptoms of COVID-19 | 19 (16.4) |
| To control the spread of the COVID-19 virus for oneself | 15 (12.9) |
| For one's own health | 12 (10.3) |
| For safety | 7 (6.0) |
| For society | 5 (4.3) |
| To protect the frail population | 2 (1.7) |
| Out of fear | 1 (0.9) |

**Table 5.** Overview of the main reasons to use the tracing app (n=147).

| Reasons | Value, n (%) |
|---|---|
| To control the spread of the COVID-19 virus in general | 45 (30.6) |
| More insight into the spread and symptoms of COVID-19 | 34 (23.1) |
| For one's own health | 19 (12.9) |
| For safety | 17 (11.6) |
| To control the spread of the COVID-19 virus for oneself | 15 (10.2) |
| For society | 9 (6.1) |
| To protect the frail population | 6 (4.1) |
| Out of fear | 2 (1.4) |

## Main Reason Not to Use the Mobile Apps

An overview of the reasons to not use the mobile apps is presented in Tables 6 and 7. For both mobile apps, privacy was mentioned as the main reason (symptom app=56.6% [64/113] and tracing app=64.8% [92/142]) to not use the mobile apps. Other reasons for not using the mobile apps were the expected usefulness of the app (symptom app=23.9% [27/113] and tracing app=13.4% [19/142]) and a fear of becoming over aware of the situation and its potential consequences, leading to unnecessary stress (symptom app=8.0% [9/113] and tracing app=11.3% [16/142]).

**Table 6.** Overview of the main reasons not to use the symptom app (n=113).

| Reason | Value, n (%) |
|---|---|
| Privacy/not willing to share information with government | 64 (56.6) |
| Doubting usefulness | 27 (23.9) |
| Over awareness/stress | 9 (8.0) |
| Doubting ease of use | 5 (4.4) |
| Doubting security | 5 (4.4) |
| No (compatible) phone | 2 (1.8) |
| The fear the use of the app will be forced by government | 1 (0.9) |



XSL•FO
RenderX

**Table 7.** Overview of the main reasons not to use the tracing app (n=142).

| Reason | Value, n (%) |
|---|---|
| Privacy/not willing to share information with government | 92 (64.8) |
| Doubting usefulness | 19 (13.4) |
| Over awareness/stress | 16 (11.3) |
| No (compatible) phone | 6 (4.2) |
| Doubting security | 3 (2.1) |
| Doubting ease of use | 3 (2.1) |
| The fear the use of the app will be forces by government | 3 (2.1) |

## Discussion

The aim of this paper was to identify antecedents of acceptance of (1) a mobile app for COVID-19 symptom recognition and monitoring, and (2) a mobile app for contact tracing among Dutch citizens by means of an online survey.

### Principal Results

Our main finding is that for both mobile apps age, attitude toward technology, and fear of COVID-19 are antecedents of acceptance. A large group of the Dutch citizens (101/223, 45.3%) is willing to use a mobile app for COVID-19 symptom recognition and monitoring. The main reasons to use this mobile app are to (1) control the spread of COVID-19; (2) monitor their own complaints, and (3) gain more insight into the spread and symptoms of the COVID-19 virus. For the case of a mobile app for COVID-19 contact tracing, 41.2% (98/238) of the Dutch adults appear to be willing to use this mobile app. The main reasons for use are (1) to control the spread of the COVID-19 virus, (2) to gain more insight into the spread and symptoms of the COVID-19 virus, and (3) for their own health. Privacy, doubting the usefulness of the mobile app, and a fear of becoming over aware of the situation and its potential consequences, leading to unnecessary stress, are the main reasons to not use the mobile apps. Overall, Dutch citizens were more willing to use a mobile app for COVID-19 symptom recognition and monitoring compared with a mobile app for contact tracing.

### Comparison With Prior Work

It is difficult to relate our findings to the existing literature, as limited technology acceptance studies have focused on mobile apps to be used during a pandemic, and insights into factors that determine the acceptance of COVID-19–related mobile apps are lacking [7]. In general, age and attitude toward technology are widely acknowledged antecedents of acceptance. For age there is evidence that older age is associated with a lower level of acceptance of mobile apps [13]. Previous results also indicated that attitude toward technology is an important antecedent of acceptance of mobile apps [13,14]. The degree to which an individual is willing to try out any new mobile app is related to the intention to use [13]. Since this study, the mobile apps, announced by the Dutch Government in April 2020, have been developed and implemented. In a recent study by Bente et al [15], the contact tracing app (the CoronaMelder) was tested for usability, and was found easy to use. A comparable study was executed in Germany by Blom et al [16]. They analyzed the potential barriers for the large-scale adoption of the official contact tracing app that was introduced in Germany. The foremost barrier toward using the contact tracing app was the lack of willingness to correctly adopt the app. Besides, compared with the younger group (aged 18-59 years), the older age group (aged 60-77 years) was less likely to use a compatible smartphone. Therefore, access was also mentioned as barrier in this study [16]. Another cross-country survey study (participating countries: France, Germany, Italy, the United Kingdom, and the United States) on the acceptance of a contact tracing app is more optimistic [17], as the willingness to install the app was high among all 5 counties and across all subgroups of the population. In addition, this study concluded that epidemiological evidence shows that app-based contact tracing can suppress the spread of COVID-19 if a high enough proportion of the population uses the app [17].

Our results show that fear of COVID-19 is the most important COVID-19–related factor that predicts acceptance of mobile apps to deal with the COVID-19 pandemic. Because it is difficult to translate this fear into a technology design, this finding needs to be seen in a bigger picture. Public health campaigns during the COVID-19 epidemic will need to educate citizens about the dangers of COVID-19 (personally and for society as a whole), and should then offer downloading COVID-19 mobile apps as a personal strategy to deal with this fear. Next, the positive attitude toward technology that precedes a decision to download a COVID-19 app should be taken into consideration when using these innovations. The end-user population might be skewed toward those with interest in technology (traditionally these are younger, highly educated men [18]), which can create a use divide, and thus, a health divide in society. Measures should be installed to support those groups in society that are not, by nature, technically interested, such as having promotional stalls in the community and diverse channels of user support.

### Limitations

The following 4 limitations should be taken into account for this study. First, due to our recruitment method (snowball sampling via social media), our sample could have been affected by a selection bias. Our sample was mainly composed of participants with a high educational level and a moderate attitude toward technology. Therefore, our results are based on the views of a somewhat skewed sample of the Dutch population, which might reduce the generalizability of our findings. Second, for



XSL•FO
RenderX

our analysis, the power of our sample was sufficient. However, a larger sample would improve the generalizability of our outcomes, as mainly Dutch citizens from the eastern part of the Netherlands (87.0% [207/238] of our sample) completed our survey. Third, in our survey the 2 mobile apps are introduced by means of a short description of their general aim. It is unclear if this description was sufficient for the responders to understand the purpose of both mobile apps. Our survey was distributed before the development of the CoronaMelder app in the Netherlands. The study by Bente et al [15] indicated that during this period there were many misconceptions concerning contact tracing among the Dutch population. It is likely that these short descriptions of the general aim of the 2 mobile apps were insufficient to take those misconceptions away. Fourth, the explained variance of both our models is relatively low. Normally, in studies such as these, this number is boosted by including the predictors perceived ease of use and perceived usefulness. However, including these 2 factors leads to little practical results, that is, concluding that the apps should be easy to use. By contrast, the identification of COVID-19–related factors remains an important extension of the existing technology acceptance models.

## Conclusions

Age, attitude toward technology, and fear of COVID-19 are important predictors of the acceptance of COVID-19 mobile apps for symptom recognition and monitoring and for contact tracing. These predictors should be taken into account during the development and implementation of these mobile apps to secure acceptance.

## Authors' Contributions

The survey was developed by SJ-K, MH, and LvV. Statistical analyses were performed by SJ-K and LvV. All authors were involved in the distribution of the survey and participated in drafting the article and revising it critically for important intellectual content.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questions and answer options in Dutch and English (D=demographic questions; C=fear of COVID-19 questions; H=perceived health questions; TAM-BI=behavioural intention).
[DOCX File , 21 KB - formative_v5i12e28416_app1.docx ]

## References


1. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? The Lancet Respiratory Medicine 2020 Apr;8(4):e21. [doi: 10.1016/s2213-2600(20)30116-8]
2. Roux J, Massonnaud C, Crépey P. COVID-19: One-month impact of the French lockdown on the epidemic burden. medRxiv 2020 Apr 27;1:1 [FREE Full text] [doi: 10.1101/2020.04.22.20075705]
3. Di Domenico L, Pullano G, Sabbatini CE, Boëlle PY, Colizza V. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. BMC Med 2020 Jul 30;18(1):240 [FREE Full text] [doi: 10.1186/s12916-020-01698-4] [Medline: 32727547]
4. Krausz M, Westenberg J, Vigo D, Spence R, Ramsey D. Emergency Response to COVID-19 in Canada: Platform Development and Implementation for eHealth in Crisis Management. JMIR Public Health Surveill 2020 May 15;6(2):e18995 [FREE Full text] [doi: 10.2196/18995] [Medline: 32401218]
5. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):722-727 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
6. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective Configurations of a Digital Contact Tracing App: A Report to NHSX. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf [accessed 2020-04-16]
7. Thorneloe R, Epton T, Fynn W, Daly M, Stanulewicz N, Kassianos A, et al. Scoping review of mobile phone app uptake and engagement to inform digital contact tracing tools for COVID-19. PsyArXiv 2020 Apr 30;1:1 [FREE Full text] [doi: 10.31234/osf.io/qe9b6]
8. Anderson K, Burford O, Emmerton L. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. PLoS One 2016 May 23;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
9. Agarwal R, Prasad J. A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. Information Systems Research 1998 Jun;9(2):204-215. [doi: 10.1287/isre.9.2.204]
10. van Velsen L, Beaujean D, van Gemert-Pijnen JE, van Steenbergen JE, Timen A. Public knowledge and preventive behavior during a large-scale Salmonella outbreak: results from an online survey in the Netherlands. BMC Public Health 2014 Jan 31;14:100 [FREE Full text] [doi: 10.1186/1471-2458-14-100] [Medline: 24479614]




XSL•FO
RenderX

11. van Velsen L, van der Geest T, van de Wijngaert L, van den Berg S, Steehouder M. Personalization has a Price, Controllability is the Currency: Predictors for the Intention to use Personalized eGovernment Websites. Journal of Organizational Computing and Electronic Commerce 2015 Feb 05;25(1):76-97. [doi: 10.1080/10919392.2015.990782]
12. CBS Open Data. URL: https://opendata.cbs.nl [accessed 2021-11-02]
13. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 02;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
14. Hussein Z, Oon S, Fikry A. Consumer Attitude: Does It Influencing the Intention to Use mHealth? Procedia Computer Science 2017;105:340-344. [doi: 10.1016/j.procs.2017.01.231]
15. Bente B, van 't Klooster JWJR, Schreijer M, Berkemeier L, van Gend JE, Slijkhuis P, et al. The Dutch COVID-19 Contact Tracing App (the CoronaMelder): Usability Study. JMIR Form Res 2021 Mar 26;5(3):e27882 [FREE Full text] [doi: 10.2196/27882] [Medline: 33724198]
16. Blom AG, Wenz A, Cornesse C, Rettig T, Fikel M, Friedel S, et al. Barriers to the Large-Scale Adoption of a COVID-19 Contact Tracing App in Germany: Survey Study. J Med Internet Res 2021 Mar 02;23(3):e23362 [FREE Full text] [doi: 10.2196/23362] [Medline: 33577466]
17. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
18. Rojas-Méndez JI, Parasuraman A, Papadopoulos N. Demographics, attitudes, and technology readiness. MIP 2017 Feb 06;35(1):18-39. [doi: 10.1108/mip-08-2015-0163]

*Edited by G Eysenbach; submitted 04.03.21; peer-reviewed by JW Klooster, P Kanzow, A Roy, J Marconato, T Capote; comments to author 25.03.21; revised version received 04.05.21; accepted 16.11.21; published 20.12.21.*

*Please cite as:*
*Jansen-Kosterink S, Hurmuz M, den Ouden M, van Velsen L*
*Predictors to Use Mobile Apps for Monitoring COVID-19 Symptoms and Contact Tracing: Survey Among Dutch Citizens*
*JMIR Form Res 2021;5(12):e28416*
*URL: https://formative.jmir.org/2021/12/e28416*
*doi:10.2196/28416*
*PMID:34818210*

©Stephanie Jansen-Kosterink, Marian Hurmuz, Marjolein den Ouden, Lex van Velsen. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# The Plebeian Algorithm: A Democratic Approach to Censorship and Moderation

Benjamin Fedoruk[1*]; Harrison Nelson[2*]; Russell Frost[3*]; Kai Fucile Ladouceur[4*]

[1]Faculty of Science, University of Ontario, Institute of Technology, Oshawa, ON, Canada
[2]Faculty of Health Sciences, Queen's University, Kingston, ON, Canada
[3]Faculty of Engineering, Lakehead University, Thunder Bay, ON, Canada
[4]School of Engineering Technology, Trades, and Aviation, Confederation College, Thunder Bay, ON, Canada
[*]all authors contributed equally

**Corresponding Author:**
Benjamin Fedoruk
Faculty of Science
University of Ontario
Institute of Technology
2000 Simcoe St N
Oshawa, ON, L1G 0C5
Canada
Phone: 1 905 721 8668
Email: benjamin.fedoruk@ontariotechu.net

## Abstract

**Background:** The infodemic created by the COVID-19 pandemic has created several societal issues, including a rise in distrust between the public and health experts, and even a refusal of some to accept vaccination; some sources suggest that 1 in 4 Americans will refuse the vaccine. This social concern can be traced to the level of digitization today, particularly in the form of social media.

**Objective:** The goal of the research is to determine an optimal social media algorithm, one which is able to reduce the number of cases of misinformation and which also ensures that certain individual freedoms (eg, the freedom of expression) are maintained. After performing the analysis described herein, an algorithm was abstracted. The discovery of a set of abstract aspects of an optimal social media algorithm was the purpose of the study.

**Methods:** As social media was the most significant contributing factor to the spread of misinformation, the team decided to examine infodemiology across various text-based platforms (Twitter, 4chan, Reddit, Parler, Facebook, and YouTube). This was done by using sentiment analysis to compare general posts with key terms flagged as misinformation (all of which concern COVID-19) to determine their verity. In gathering the data sets, both application programming interfaces (installed using Python's pip) and pre-existing data compiled by standard scientific third parties were used.

**Results:** The sentiment can be described using bimodal distributions for each platform, with a positive and negative peak, as well as a skewness. It was found that in some cases, misinforming posts can have up to 92.5% more negative sentiment skew compared to accurate posts.

**Conclusions:** From this, the novel Plebeian Algorithm is proposed, which uses sentiment analysis and post popularity as metrics to flag a post as misinformation. This algorithm diverges from that of the status quo, as the Plebeian Algorithm uses a democratic process to detect and remove misinformation. A method was constructed in which content deemed as misinformation to be removed from the platform is determined by a randomly selected jury of anonymous users. This not only prevents these types of infodemics but also guarantees a more democratic way of using social media that is beneficial for repairing social trust and encouraging the public's evidence-informed decision-making.

***(JMIR Form Res 2021;5(12):e32427)*** doi:10.2196/32427

**KEYWORDS**

infodemiology; misinformation; algorithm; social media; plebeian; natural language processing; sentiment analysis; sentiment; trust; decision-making; COVID-19



XSL•FO
RenderX

## Introduction

The internet is a powerful tool for spreading information; as such, it follows that it is equally powerful for spreading misinformation. In 2019, the number of social media users worldwide was 3.484 billion [1], with that number increasing year-by-year by an average of 9% [1,2]. With this increased use, the "power-user" or microinfluencer phenomenon has arisen, where popular social media accounts are able to reach large numbers of readers. This is increasingly important as more people begin to use social media as a source for news [3]. This news comes from a third party by a popular influencer, not posted or moderated by the social media companies themselves. Past analyses examining online misinformation often classify posts as misinformation using a "Point-And-Shoot" algorithm; this is the status quo. However, some algorithms will be better at combating misinformation than others. The Plebeian Algorithm creates criteria that social media websites should take into account when designing their algorithms to reduce misinformation. This reduction of misinformation is thought to be achieved by examining the correlation between sentiment and misinformation; it has been found that posts containing misinformation tend to have more negative sentiment when compared categorically to other posts covering the same issue [4]. Due to this correlation, it is hypothesized that an algorithm that encourages positive interactions will also reduce the amount of misinformation present on the platform through a democratic manner.

Misinformation is a key problem, yet many terms are confused in studies. Herein, the authors shall define several key terms that are often used interchangeably but whose definitions are specific and distinct. First, misinformation shall be defined as the spread, intentional or otherwise, of false information [5]. The intentions of the individuals spreading the information is irrelevant. Second, disinformation is the purposeful spread of false information [5]. A similar yet distinct definition is malinformation, which is the malicious spread of false or misleading information [5]. Finally, fake news is defined as any misinformation (with or without intention) that readers interpret as trustworthy news [5]. For this study, misinformation will be studied in-depth; however, it should be noted that future supplemental studies could conduct a similar investigation focused on disinformation, malinformation, or fake news. Infodemics can additionally be applied to the realm of health care; infodemics have the potential to intensify outbreaks when there is uncertainty among the public concerning evidence-informed preventative and protective health measures [6].

Prior to investigating the spread of misinformation, it is pertinent to define the concept of infodemiology and misinformation. This research paper defines an infodemic according to the World Health Organization (WHO) as "too much information including false or misleading information in digital and physical environments during a disease outbreak," which "causes confusion and risk-taking behaviours that can harm health [and] also leads to mistrust in health authorities and undermines the public health response" [6]. The study of the spread of infodemics on a large scale, especially pertaining to medical misinformation, is known as infodemiology. The WHO has additionally linked the rapid surge of such infodemics during the COVID-19 pandemic to "growing digitization," which can support the global reach of information but can also quickly amplify malicious or fabricated messages [6]. The second relevant definition is that of the concept of misinformation, which has been defined similarly to the definition used by the WHO in relation to the infodemic [6], but specifically refers to the distinct lack of verity in information related to a specific field.

At their core, most social media websites aim to maximize the amount of time that users spend on their platforms. This maximization of user page time leads to companies using highly specialized and trained machine learning to advertise content on users' feeds [7]. At the same time, this can have unintended adverse effects such as maximizing the time a user engages with content that is not verified for accuracy. The proposed solution to this disparity between engagement and integrity is to create democratically moderated spaces. Democratic spaces and recommendations to posts with more positive sentiment are integral concepts in the Plebeian Algorithm, based on the latest evidence that misinformation tends to be more negative [5]. The Plebeian Algorithm is an algorithm, described herein, for the purpose of the control of the spread of misinformation on social media. It is beneficial compared to other existing algorithms, as will become evident. The currently implemented point-and-shoot algorithms are hypertuned to specific sources of misinformation surrounding specific topics. However, they are not adaptable to the fluidity of the definition of true information.

As mentioned earlier, most social media platforms work on a model similar to Twitter, Facebook, or YouTube where content is recommended based on user engagement [7]; however, this is not true to the same extent for all websites. One example of a website that breaks the expectations for social media algorithms is 4chan. 4chan is an excellent epitome of ephemeral social media, where content is completely anonymous and is rapidly discarded regardless of popularity [8]; in addition, there is close to no moderation and the content tends to be more negative in sentiment. This is also exemplified in Parler, an alternative social media platform established in September 2018 that aimed to bring forth a platform with total freedom of speech. Consequently, Parler attracted those who were banned from other social media websites, creating "echo chambers, harbouring dangerous conspiracies and violent extremist groups" [9] such as those who were involved in raiding the US Capitol on January 6, 2021. Reddit also has a forum-based system similar to 4chan. However, individual fora on these platforms have moderators who work to combat negative sentiment throughout the website. Reddit's issue lies in its incredibly isolated fora, as tailoring one's feed to be a vast majority of explicitly handpicked fora is a part of the experience; this allows for some fora to have little to no moderation [10].

There have been several related works of research in the field of misinformation detection on social media platforms. These works include studies on the connection between misinformation and cognitive psychology [11], the analysis of geospatial infodemiology [12], the effect of recommendation algorithms



XSL•FO
RenderX

on infodemiology [13], the use of distributed consensus algorithms to curb the spread of misinformation [14], and the naming conventions used for viruses [15]. Although these works are in alignment with this study, they do not propose the same solution. The study that offers a solution closest to that proposed by the Plebeian Algorithm discusses the efficacy of curbing the spread of misinformation through layperson judgements [16]. Notably, this work discusses the merits surrounding a layperson algorithm but does not make suggestions for its implementation.

The objective of the study will be to determine the optimal social media algorithm to reduce the spread of misinformation while ensuring personal freedoms. The investigation conducted in this paper will have far-reaching implications that will alter how misinformation in social media platforms is addressed.

The three major implications include:

- The creation of a more open and democratic environment on social media platforms
- An overall reduction in political divisiveness and extremist sentiment both online and offline
- An increase in informed users who can make well-informed opinions on subjects

## Methods

A detailed step-based methodology was used to analyze data throughout the research process. Python 3.9 (Python Software Foundation) was the language of choice through all aspects of the project. All libraries used can be accessed using pip. The visualization of data was performed using the matplotlib and seaborn libraries in Python. Application programming interfaces (APIs) were used from Twitter, Reddit, and 4chan, gathering data regarding username, date, post, and text. Furthermore, two data sets were gathered from academic sources, containing post data from Twitter [17] and Parler [9]. Various Python libraries were used to interact and connect with the APIs, including twarc, urllib3, and basc_py4chan. The following Python libraries were used to clean the data: beautifulsoup4, demoji, and pyenchant. The pandas library for Python was used to retrieve and store third-party data sets [8,17-20], and the numpy library was used for various array operations. Finally, the nltk library was used to perform sentiment analysis, and sklearn was used to perform regressions.

Python was selected due to its ease of connectivity to the various APIs; it is well supported among a strong community, and as such, connecting to various APIs was done through prewritten libraries. This reduced the programming time while increasing the efficiency and reliability of the code.

In three of the social media services for which APIs were used (ie, Twitter, Reddit, and 4chan), four steps were performed: (1) gather data using the API and the associated Python library; (2) clean data to create a Python set of strings, containing no URLs (removed using regular expressions), HTML (removed using beautifulsoup4), usernames (removed using regular expressions), emojis (replaced with text using demoji), or non-English language (removed using pyenchant); (3) perform sentiment analysis using nltk's SentimentIntensityAnalyzer class; and (4) save the cleaned and sentiment analyzed data frame as a pickle file. A visualization script was then programmed to display the sentiment data gathered from the social media post data. To ensure confidentiality of users, only aggregate data was displayed. Plotting a histogram with a kernel density estimation (KDE) resulted in the various graphs produced by the research team. Data for which sentiment analysis returned inconclusive due to textual limitations was removed from visualization. Limiting the language to English has the benefit of statistical comparison congruence. One notable platform that did not use an API is Facebook. The reason for this is due to the restrictions placed on the Facebook API, in terms of depth and breadth of research.

The six social media services analyzed (4chan, Twitter, Parler, Reddit, YouTube, and Facebook) had various amounts of associated data. A breakdown of the data analyzed herein is described in Figure 1.

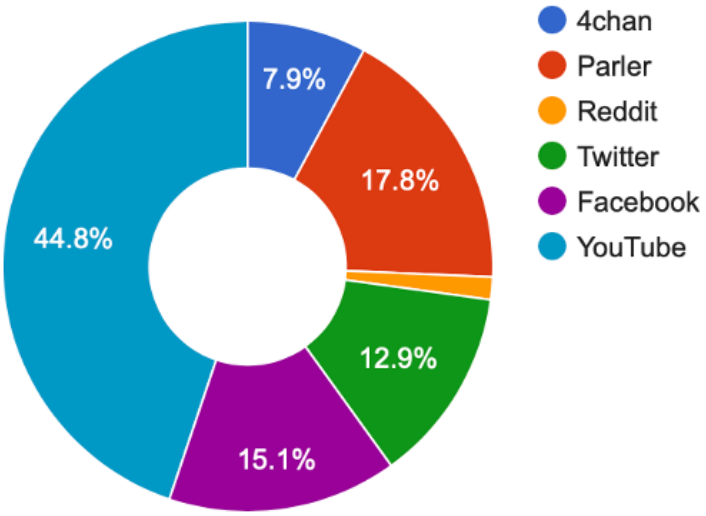
The sentiment analysis dictionary selected for the analysis performed herein is the Valence Aware Dictionary and Sentiment Reasoner (VADER). This dictionary was selected as it is the industry standard for a wide array of general statement analyses and is especially recognized for producing highly accurate results with social media platforms. As such, VADER is the optimal dictionary for the purposes of this research. Although ideal algorithms should implement various checks and balances for the sentiment analysis system implemented, this paper shall focus on strictly the VADER dictionary, which is solely positive and negative sentiment. Other sentiment analysis tools exist to examine specific emotions (including anger, fear, surprise, happiness, etc).

Key term analysis was used for the data cleaning process to determine which strings were classified as relating to a specific topic. The key terms were gathered using a list of the most commonly held terms gathered from Twitter that were directly associated with misinformation.



XSL•FO
RenderX

**Figure 1.** Social media data breakdown.




To confirm the academic literature [5] regarding the correlation between negative sentiment and verity of information, the analysis was performed using Twitter. Data was filtered such that only Tweets containing a set of potentially misinformative keywords were assigned to be assessed using a sentiment analysis. Both were plotted through histogram, and the KDEs were compared (relative to each respective maxima).

Misinformation is directly correlated to negativity. A misinformative post is often negative in sentiment. However, this is not a certainty. As such, when determining an optimal algorithm, it will be critical to use sentiment analysis to narrow the potential misinformative candidates and then to use further methods—a jury process—to accurately detect misinformation.

The study defines several mathematical terms. Many of the histograms and KDEs as previously described form a bimodal distribution. The polarity score upon which the two peaks are centered is termed $\mu^+$ and $\mu^-$, where the sign indicates whether the term refers to the positive or negative peak. The other variable defined is the skewness of the distribution as a whole, which is described using the symbol $\gamma$. When the positive peak is the major mode, then $\gamma \in (0, \infty)$. Contrarily, when the positive peak is the minor mode, then $\gamma \in (-\infty, 0)$. The frequency function $f$ describes the frequency curve represented by the KDE (such that $f(p)$ represents the frequency of strings with polarity score $p$). The skewness is calculated using the following equation:

This equation for skewness was derived using the following derivation:

## Results

The results of the analysis will be divided by the social media platform. They will be presented in the following order:

- Reddit
- 4chan
- Facebook
- YouTube
- Parler
- Twitter

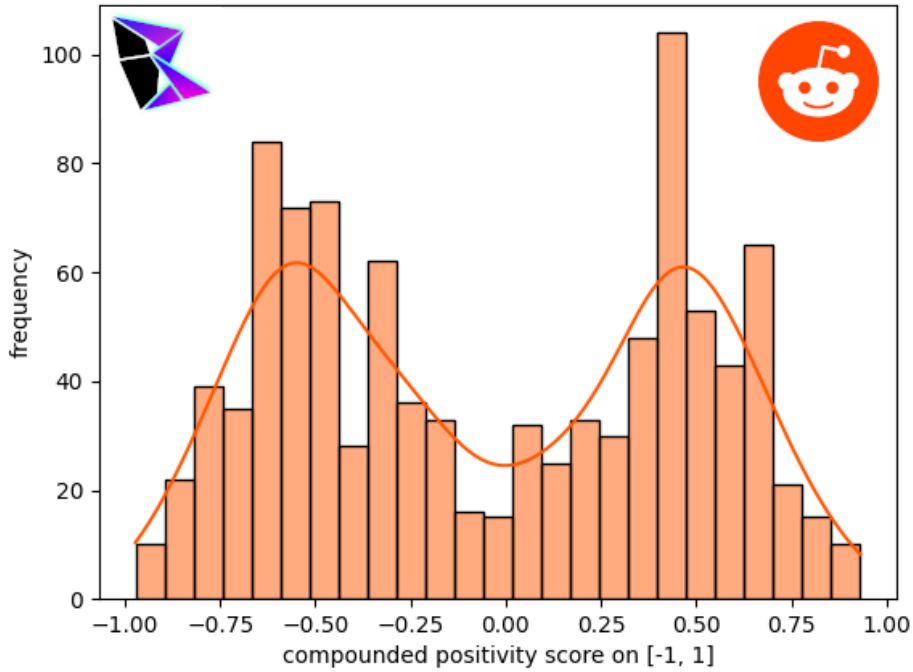
### Reddit

When analyzing Reddit's data, a series of subreddits were selected. The subreddits selected were r/AskReddit, r/AskThe_Donald, r/conspiracy, r/covid, r/kindness, r/movies, r/politics, and r/EnoughTrumpSpam. These subreddits were selected as an array of options, allowing an analysis of probable misinformative, probable truthful, and unknown sources. The data was gathered using the Python library urllib3. The first subreddit to be examined herein is r/AskReddit. This subreddit tends to contain a wide variety of posts from a myriad of conversation topics. As such, it is relatively indicative of Reddit overall. r/AskReddit's histogram can be found in Figure 2. The bimodal distribution was $\mu^- \approx -0.54$, $\mu^+ \approx 0.48$, and $\gamma \approx -0.03214$.



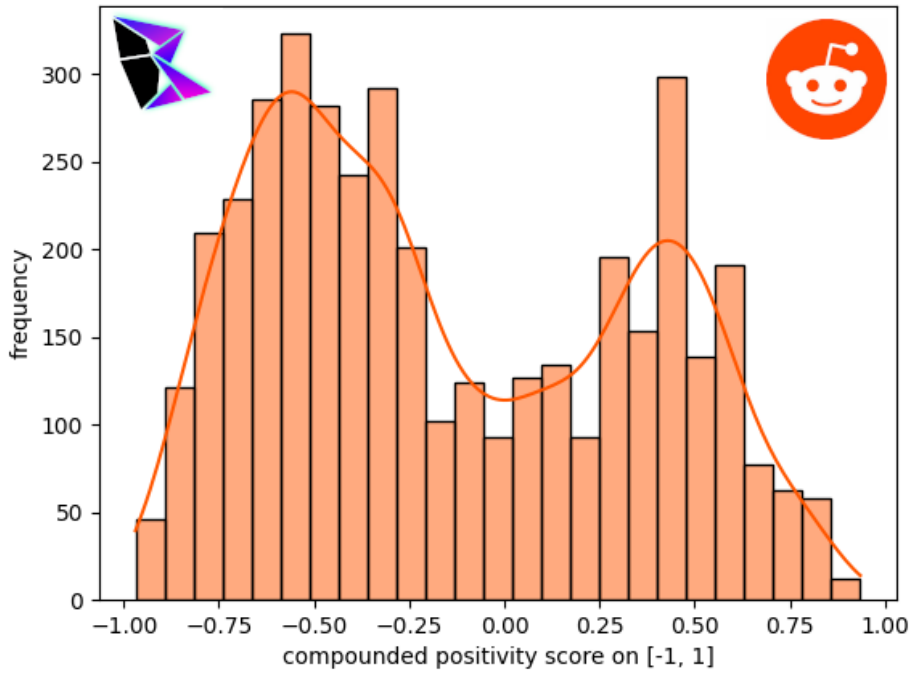
XSL•FO
RenderX

**Figure 2.** Reddit r/AskReddit frequency of positivity histogram.




It should be noted that the extremal frequencies of the bimodal distribution were approximately equal between the negative and positive peaks. Another notable subreddit examined was r/politics, which provided a sample of posts potentially swayed by the political leaning of Reddit users. The histogram and KDE for this analysis is displayed in Figure 3. The bimodal distribution for r/politics was $\mu^{-}\approx-0.56$, $\mu^{+}\approx0.43$, and $\gamma\approx-0.37776$.

**Figure 3.** Reddit r/politics frequency of positivity histogram.

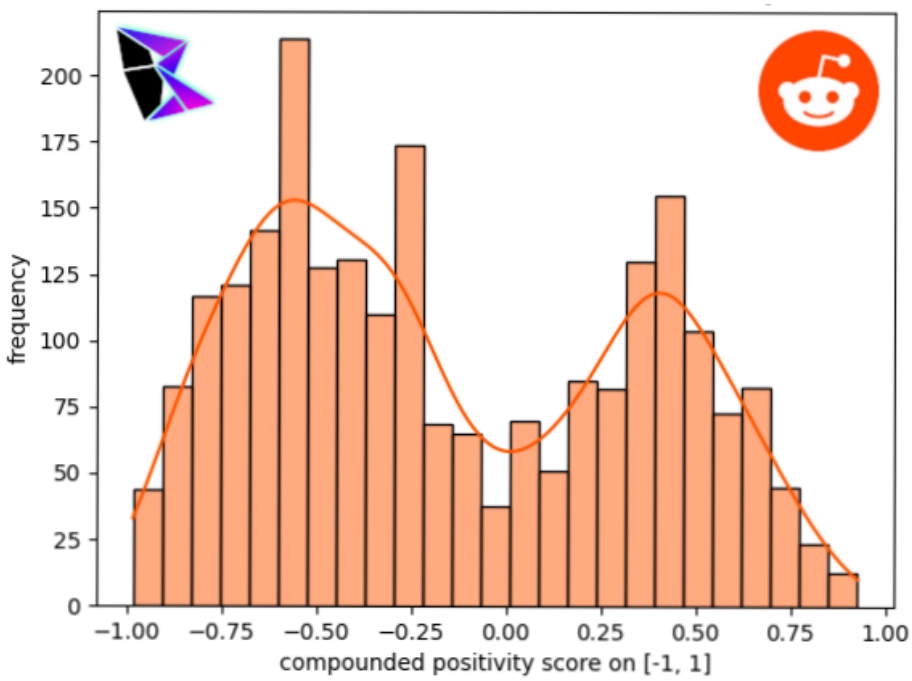



r/politics's content had a stronger negative skew as is apparent by the KDE. The final subreddit to be examined is an avant-garde subreddit: r/conspiracy. In this community, users share various conspiracy theories. When one scrolls through r/conspiracy, plenty of misinformation can easily be noted, including misinformation surrounding Flat Earth Theory and QAnon. The histogram for r/conspiracy is found in Figure 4. The bimodal distribution for r/conspiracy was $\mu^{-}\approx-0.56$, $\mu^{+}\approx0.39$, and $\gamma\approx-0.33904$.



XSL•FO
RenderX

**Figure 4.** Reddit r/conspiracy frequency of positivity histogram.

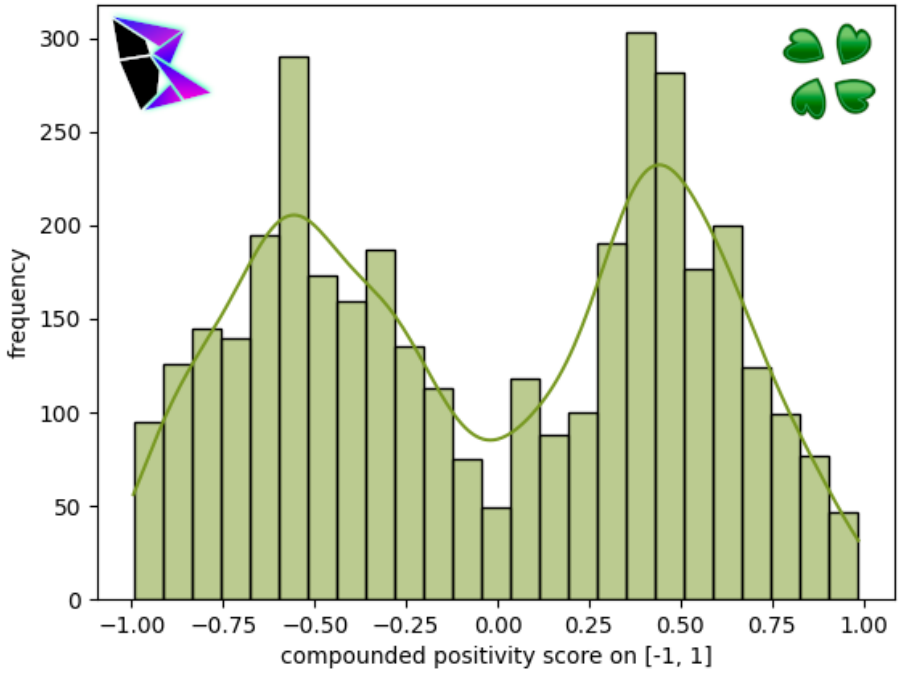



As can be noted by r/conspiracy, the conspiratorial posts (which are known to contain a large volume of misinformation) are more often negative. This can be noted due to the difference in the peaks of the bimodal distribution.

## 4chan

To analyze the 4chan data, five boards were selected: /b/, /a/, /v/, /pol/, and /r9k/. These five boards were selected due to their high post frequency compared to other 4chan boards. The data was gathered using basc_py4chan. For each of these boards, a histogram was plotted (with an overlayed KDE) with 30 bins. A visualization for the histogram for /b/'s sentiment can be found in Figure 5. /b/ is described as the random board, containing a wide mixture of conversation from across 4chan. The bimodal distribution for /b/ was $\mu^{-}\approx-0.55$, $\mu^{+}\approx0.46$, and $\gamma\approx0.11380$.

**Figure 5.** The 4chan /b/ frequency of positivity histogram.




Another board to be visualized in this report is the visualization for the sentiment of /pol/, which can be found in Figure 6; /pol/ contains political discussion. The bimodal distribution for /pol/ was $\mu^{-}\approx-0.61$, $\mu^{+}\approx0.38$, and $\gamma\approx-0.16559$.

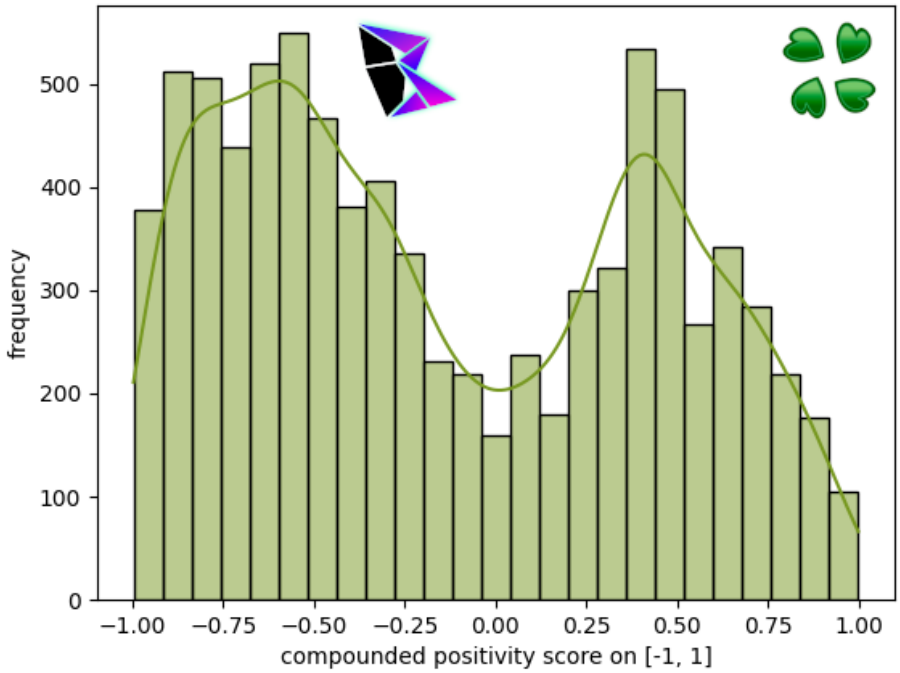
It should be noted that the levels of extreme negative sentiment (ie, with a polarization score of less than 0.75) are substantially higher in /pol/ compared to /b/. This demonstrates that political topics tend to be more negative on 4chan.



XSL•FO
RenderX

Overall, it should be noted that 4chan consistently contains a large number of negative posts, which is greatly dependent upon the topic of the board. Boards which pertain to specific recreational activities (eg, /v/ for video games or /a/ for anime) have a lesser degree of negative polarity.

**Figure 6.** The 4chan /pol/ frequency of positivity histogram.




## Facebook

It is pertinent for this paper to perform an analysis on Facebook, which is currently the social platform with the largest user base of 2.8 billion active monthly users [21]. Facebook has proven to be the social media platform with the highest user base, and as such it is pertinent for this paper to perform analysis on data collected for Facebook. A data set that specifically contains data predating the COVID-19 pandemic was accessed to broaden the scope of the sentiment analysis [22]. The data set includes data gathered from Facebook's inception until 2017. A data set with a random selection of Facebook comments from the temporal range [22] was selected for sentiment analysis using VADER.

In Figure 7, a histogram was plotted with 30 bins, depicting the frequency of Facebook comments at various sentiment analytic levels. A KDE was overlayed onto the plot to show the general trend.

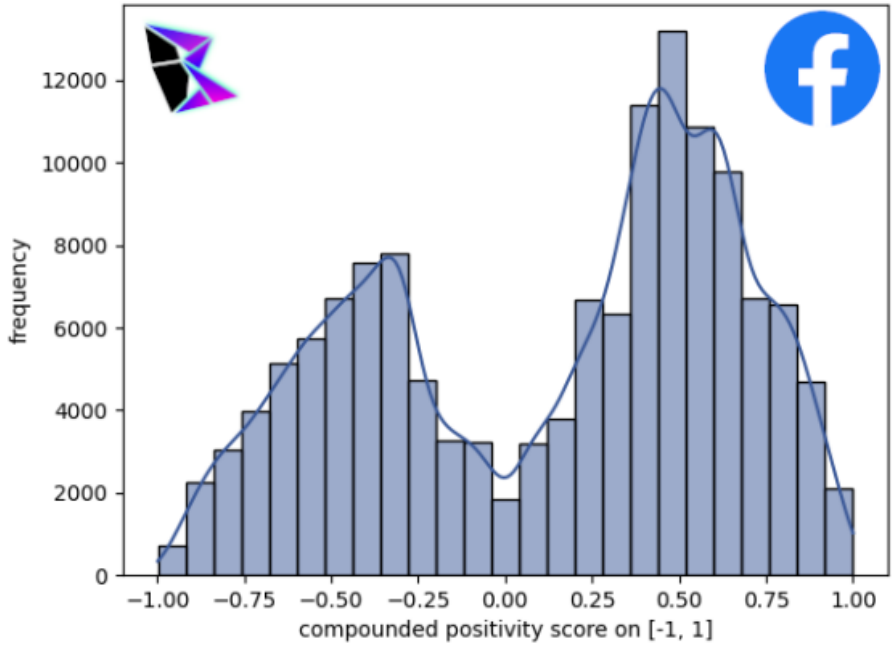
Notable features of the bimodal histogram include the sharp positive peak and wide negative peak. It should be noted that the integral for the KDE is as follows:

Furthermore, the following values were extracted from the KDE: $\mu^{+}\approx 0.43$, $\mu^{-}\approx -0.29$, and $\gamma\approx 0.49858$.



XSL•FO
RenderX

**Figure 7.** Facebook frequency of positivity histogram.



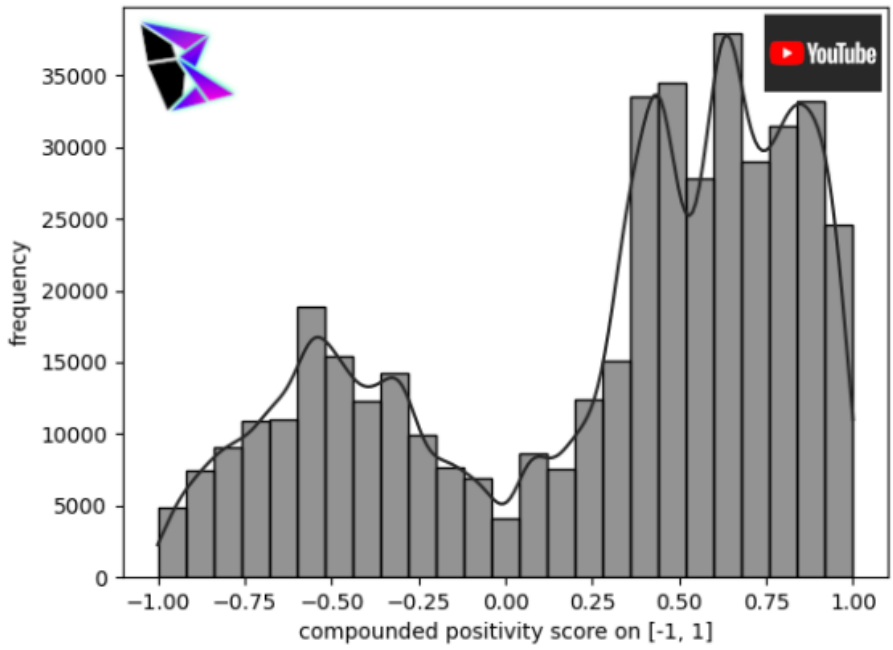

## YouTube

YouTube is based entirely on long-form video content and tends itself toward more in-depth topics. A preselected data set of YouTube comments [23], after sentiment analysis, has been visualized and presented in Figure 8.

The data set [22] was collected in 2017 and, as such, does not contain misinformation related to COVID-19. This helps to broaden the temporal scope of this analysis and ensure that the present trends hold in data outside of the COVID-19 pandemic (ie, prior to January 2020). It was also limited in geographic scope to the United States, the United Kingdom, and Canada. This limitation was due to the availability of data. It should be noted that these three countries represent the English-speaking members of the Group of Seven, a group of the seven most democratic, affluent, and pluralist nations in the world.

As can be noted, there was a strong positive skewness in the data, with $\mu^{+}\approx 0.67$, $\mu^{-}\approx -0.52$, and $\gamma\approx 0.66593$. The high positive skewness should be noted for these YouTube comments. Potential explanations for this trend will be discussed in a later section.

**Figure 8.** YouTube COVID-19 frequency of positivity histogram.




## Parler

The analysis of Parler was a transition from the traditional analyses of Reddit and 4chan, due to the fact that Parler is not broken down into communities to which users subscribe but is a single news feed–style system. The analysis of Parler should be contrasted to the analysis of Twitter in the subsequent section,

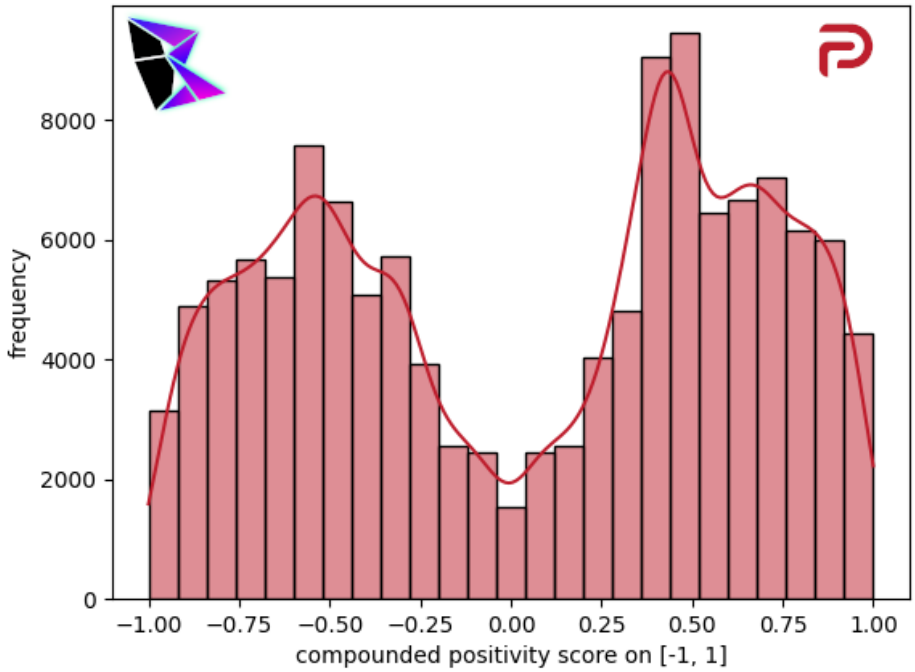


XSL•FO
RenderX

as users migrated from Twitter to Parler due to a perception of limitations placed on their freedom of expression on Twitter.

When analyzing Parler, data was collected into a data set throughout the COVID-19 pandemic and the period surrounding the events of January 6, 2021 [9]. Figure 9 contains a visualization of the COVID-19–related parleys posted between January 2020 and March 2020. The bimodal distribution was $\mu^{-}\approx-0.53$, $\mu^{+}\approx0.45$, and $\gamma\approx0.22063$.

**Figure 9.** Parler COVID-19 frequency of positivity histogram.




## Twitter

The majority of the analysis performed through this paper was on the social media service Twitter. The reason for this is due to the high amount of data regarding misinformation on the platform, the overall popularity of the platform as a general case study, and the generality of the platform (compared to some other unorthodox data sources such as YouTube comments).

Similar to Parler, Twitter tweets are made at-large to the public. There are no channels, boards, or subreddits of any sort. However, due to the Twitter algorithm, there is an allowance of individuals' feeds to be in an echo chamber. Evidently, echo chambers should be avoided wherever possible. Echo chambers are a large contributor to the rampant spread of misinformation that is seen surrounding the COVID-19 pandemic [24].

This study used a combination of both data gathered from the Twitter API and a data set of pregathered COVID-19 tweets [17,18,20]. The interface used to connect the Python code (and sentiment analysis) to the Twitter API was twarc. The reason for this duplication of analysis was to ensure that the data used was accurate. Precision must be maintained in both data collected by APIs and over a long duration.

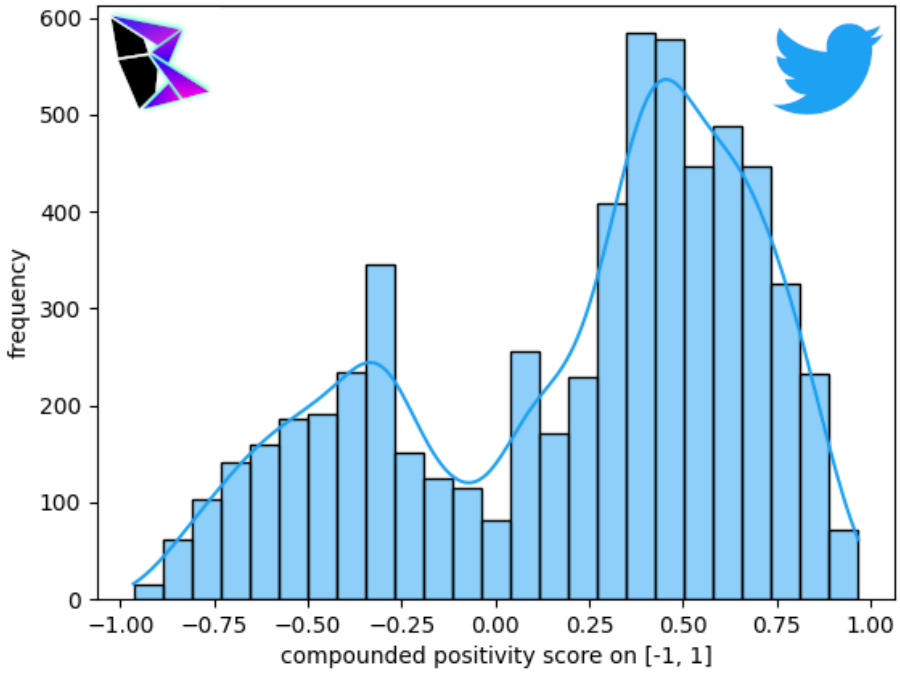
In both studies (using the API and the data set [17]), the study analyzed the broad sentiment of COVID-19–related tweets and filtered the data by keyword. The keywords used included terms concerning misinformation surrounding COVID-19, including "China Virus," "Bioweapon," and "Microchip." The filtered data then underwent sentiment analysis. Both sentimentally analyzed data were plotted on the standard histogram with overlaid KDE.

For the discussion, the study focused on the data gathered from the Twitter API, since a similar methodology was used for gathering data for the other social media platforms studied. However, it should be noted that similar results were attained using the data set [17]. Figure 10 is a graph of the Twitter API's gathered tweets pertaining to COVID-19 (the broad topic), where the sentiments of the tweets are plotted on a histogram with a KDE. A random sample of the data was taken for this analysis, as there were too many tweets to reasonably analyze the population. The bimodal distribution was $\mu^{-}\approx-0.36$, $\mu^{+}\approx0.47$, and $\gamma\approx0.86500$.



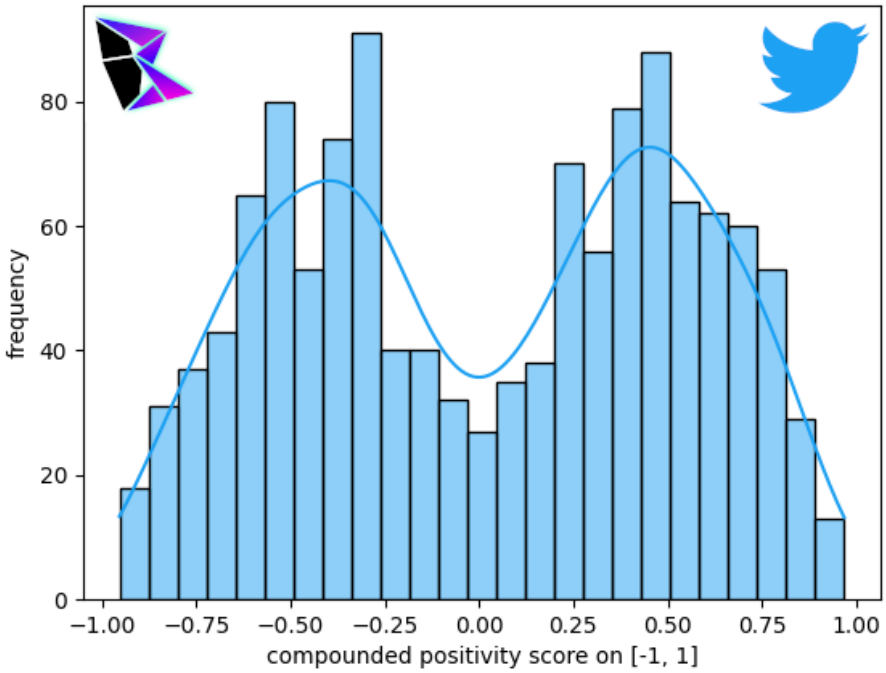
XSL•FO
RenderX

**Figure 10.** Unfiltered COVID-19 Twitter application programming interface frequency of positivity histogram.




As can be noted, the positive peak for the KDE is nearly double the negative peak. This indicates that the number of positive tweets far exceeds the number of negative tweets. Comparatively in Figure 11, the negative peak of the bimodal distribution is on par with the positive peak. This figure is a histographic representation of the polarity score for tweets after being filtered. The tweets selected only contain terms that are known to pertain to COVID-19 misinformation. The bimodal distribution was $\mu^{-}\approx-0.42$, $\mu^{+}\approx0.47$, and $\gamma\approx0.80080$ between the two Twitter measurements.

**Figure 11.** Filtered COVID-19 Twitter API frequency of positivity histogram. API: application programming interface.




The study's discussion of the reasoning behind this proportional increase in the negative peak compared to the positive peak will be discussed further in the subsequent section. Again, it must be noted that the same results can be seen when performing an identical analysis on the data gathered from the data set [17].

## Discussion

In the Discussion section, not only will an analysis of the results and errors be explored but also the Plebeian Algorithm and its benefits will be discussed, as well as how it compares to the algorithms of the social media platforms studied.



XSL•FO
RenderX

## Results Analysis

Several critical notes must be made with regard to the analysis of the quantitative features produced in the Results section.

First, it is critical to note that 4chan was the only social media platform studied that had an overall positive $\gamma$, notably on the all-encompassing board of /b/. It was gathered by the observations that a system that provided users with the freedom to determine which content got promoted—as opposed to an artificial intelligence algorithm—improved the sentiment of the average post. This is a key point in the Plebeian Algorithm, which is described in a subsequent section. Second, Twitter had a more moderate skew (ie, closer to 0, or neutral) $\bar{\mu}$, indicating that users tended to be more positive than users on the other social media platforms analyzed in the study.

It is also critical to recall that there was a strong correlation between polarity scores as determined by a sentiment analysis algorithm and the verity of the information communicated [5]. As such, the analysis provided herein can be applied to both the sentiment and the verity of a social media post.

## Echo Chambers

In analyzing the data represented by the KDEs and the skew of the bimodal distribution of the data toward negative sentiment, a confirmation of the echo chamber effect (a theory that states "the Internet has produced sets of isolated homogeneous echo chambers, where similar opinions reinforce each other and lead to attitude polarization" [25]) was clearly shown, as negative sentiment has a clear association with emotions such as anger, which have been shown to "...[reinforce] echo chamber dynamics...in the digital public sphere" [25]. In fact, other studies have also predicted the link to this effect to be the impact of the specific algorithms used in the virtual space [26]. The link gives strong evidence to suggest that the algorithms currently deployed by social media companies are creating the optimal medium through which misinformed opinions and content can grow and go uncontested. These echo chambers ensure that users are unable to get access to arguments that conflict with their beliefs and expand their perspectives.

## Sources of Error

Although the study attempted to limit error, there remained several sources of error stemming from the methodology of analysis used. The first source of error is the trouble with using key term searching, as it would give us results of posts of not only individuals spreading misinformation but also those trying to bring attention to the issue of misinformation and those who spread it. Furthermore, the sentiment analysis would also have been unable to differentiate between a misinformed post and one that tried to bring attention to the problem. This is because some of the true tweets demonstrate overall negative emotions. The final problem with the key term search is that some of the key terms determined to be misinformative may in the future be proven to be accurate information.

A further source of potential error comes in the form of the social media platforms used. One problem is that only four social media platforms were assessed, thus limiting the scope of the study. It could be that the trends found in the research will not show up in other social media platforms (eg, Facebook). Another issue from the sources assessed was that they were all only textually based and thus the method proposed may not be replicable for more graphically based social media platforms (eg, YouTube, Instagram, or Snapchat).

## Definition and Implementation of the Plebeian Algorithm

As described previously, the Plebeian Algorithm is a novel algorithm for identifying and removing misinformation through democratic means. It works in two distinct phases: the Flag Phase, to determine which posts are misinformative, and the Jury Phase, to judge the information to determine if removal is appropriate.

### *Flag Phase*

The Flag Phase is tasked with the determination of possible misinformed posts. In doing so, the algorithm selects posts that have a large number of views and then performs sentiment analysis on both the original post and a "without-replacement simple random sample" [27] of comments or replies to the post. If the overall sentiment leans negative, then the post is flagged as being potentially misleading. The posts flagged will then be passed into the Jury Phase.

### *Jury Phase*

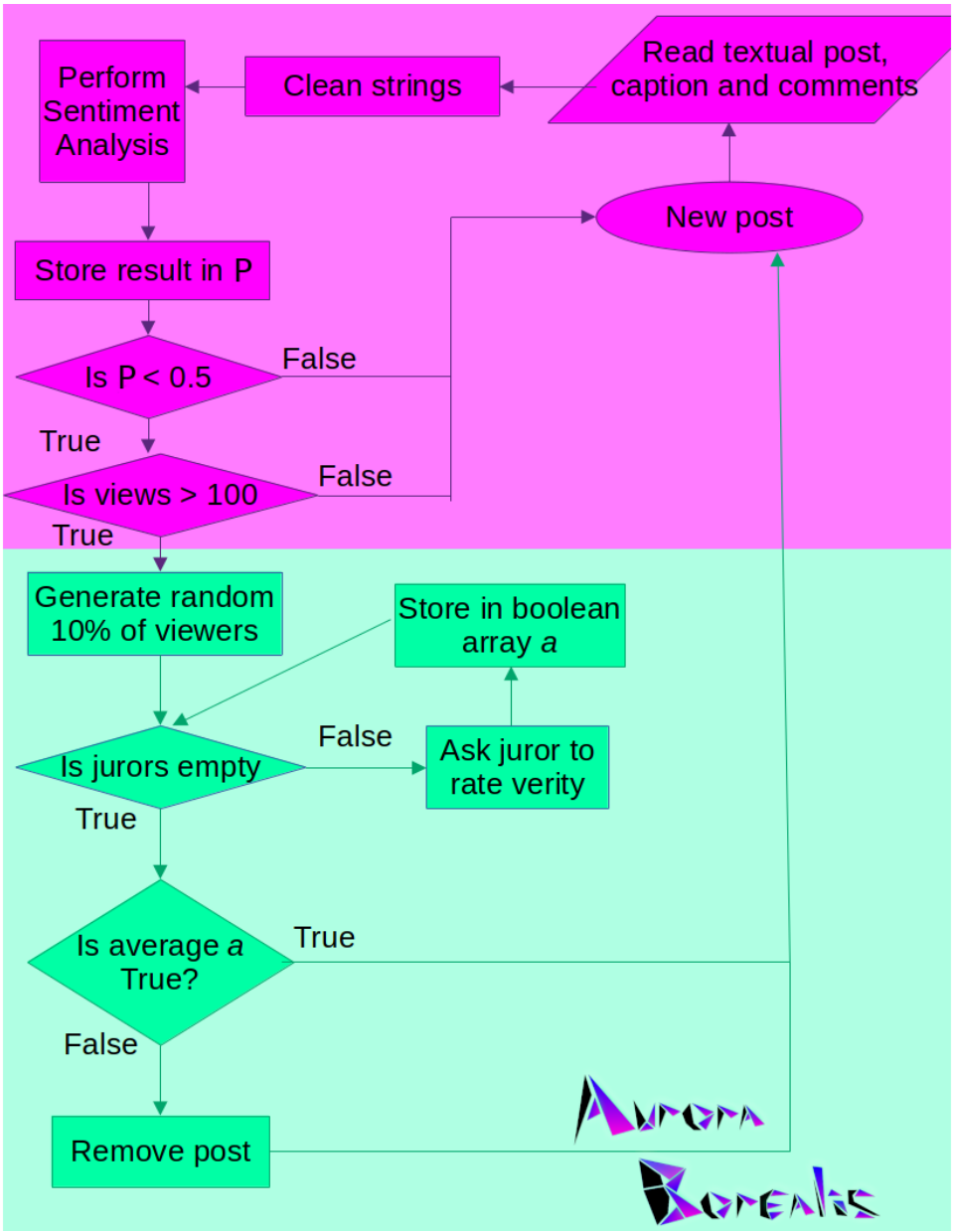
The Jury Phase is tasked with the trial and removal of truly misinformative posts. These posts are removed from the home page or news feed of the user. During the Jury Phase, flagged posts are sent to a random selection of anonymous users known as jurors. This selection should provide a diverse group, consisting of varying political opinions to give the post a fair trial. The selection of jurors uses a "without-replacement simple random sample" of a population [27]. The number of jurors selected is exactly 10% of the number of viewers of a post, rounded up. It should be noted, however, that jurors are not forced to participate or vote. It is assumed that the number of voting jurors will be far less than the total number of jurors. Thus, selecting 10% of the population allows room for the uncertainty of juror engagement. The jurors are then asked to vote either for or against the removal of the post. Once the deliberation has lasted a set duration (or a threshold of response has been met), the results will be counted and the post will remain on the site or be removed by the algorithm.

For reference, a summative flowchart detailing the Plebeian Algorithm as described can be found in Figure 12. The areas colored in magenta constitute the Flag Phase, while the areas colored in mint green constitute the Jury Phase.



XSL•FO
RenderX

**Figure 12.** Flowchart of the Plebeian Algorithm.




## Existing Algorithms

In the following section, each of the algorithms will be detailed for Reddit, 4chan, Parler, and Twitter. These analyses will be based on academic journal articles [10-12,28,29]. It is pertinent to analyze these on a case-by-case basis, as it must be ensured that the user base remains loyal to the brand and platform [30]. Prevalent existing algorithms include the PageRank and Hits algorithms [31].

Before dissecting the individual social media algorithms that are currently being used for the various platforms, it is critical to mention that most of these algorithms have an identical goal: to get users to stay on the platform, thus ensuring a continued revenue for their organization. This objective contradicts the goal of preventing misinformation from spreading on the platform, as preventing misinformation requires some censorship, resulting in a reduction of revenue. This, however, is not to indicate that the Plebeian Algorithm is of little value to site entrepreneurs. It is critical to give note that the ultimate goal of social media companies (with the exception of Parler) have already shown themselves interested in curbing and moderating their own social media platforms through their implementation of “point-and-shoot” algorithms as well as censorship of high-profile posts and accounts (eg, Twitter banning @realDonaldTrump). However, as has already been described, these algorithms are not effective at accomplishing their mission of reducing misinformation and, further, have caused users to become disillusioned with the service. This has led to many users joining platforms that capitalize on this disillusionment (eg, Parler). Hence, by implementing the Plebeian Algorithm, these social media companies finally have a method that carefully balances moderation with freedom of expression that will reinspire a sense of awe within their user base and bring back the notion of social media being a fun online space where people can collaborate and share freely.



XSL•FO
RenderX

### Reddit

The algorithm used for Reddit is a simple upvote/downvote system, as was described in the Introduction section. Users of Reddit are encouraged to upvote content they like and are encouraged to downvote content that they do not like. Posts with more upvotes are more widely shared, whereas the opposite is true with posts with more downvotes. In Reddit, users are allowed to vote on the original post and any comments. "Comment trees" are inherently created by the system as users comment on comments (thereby chaining comments together into a treelike formation).

The Reddit algorithm is tailored to the interests of the Reddit user. Through a system of subscriptions to various topics of conversations or subreddits. Users will receive a mixture of content from the subreddits to which they have subscribed, with additional, sporadic advertisement.

The system is essentially tailored to the specific user. This contrasts with the Plebeian Algorithm, which emphasizes the democratic process for the determination of verity by the user base. Currently, Reddit contains no user-controlled means to fight misinformation aside from the "Report" button, which brings the issue to the attention of a staff person at Reddit. This process is considered a manual review by the corporation, and as such, it does not constitute something similar to the Plebeian Algorithm. For Reddit to implement a Plebeian Algorithm, it must ensure that the process of the misinformation determination remains in the hands of the user base.

Although Reddit currently appears to be a democratic system, it is more of a fiefdom [29]. For example, in 2013 the r/FindBostonBombers subreddit slandered the Brown family by connecting them with the 2013 attack on the Boston Marathon at the direction of the moderators of the subreddit [32]. Examples like this resonate throughout Reddit through incidents such as "the Fappening," where nude photographs were released to the public unbeknown to victims. Incidents like these make apparent the crux of the fundamental issue with Reddit: the moderators. This promotes content moderation by a few elite members of communities, instead of by the members of the said community as a whole.

It should be noted that Reddit is a platform that is built on a sense of anonymity. Users are not required to add their personal email addresses or their real names. As is the case in all the implementations of the Plebeian Algorithm, it is important that the social media company critically analyzes the existing market served and existing qualities that users may be drawn to. An implementation of the Plebeian Algorithm on Reddit should preserve user anonymity and should still not require the use of personal emails or real full names.

### 4chan

The 4chan algorithm is similar to that of Reddit; it uses a system whereby the audience determines whether or not content is viewable to its users. In contrast to Reddit, 4chan uses an ephemeral system for its content [10]. 4chan is also divided into several boards that encapsulate distinct topics of conversation. Furthermore, any content can be posted on the /b/ board, as this board's topic is described as "Random" [10]. A second critical aspect of the 4chan algorithm is the notion of anonymity. 4chan encourages its user base to remain anonymous through their posts. Over 90% of posts and comments on /b/, the most popular board on 4chan, are anonymous [10].

4chan is the algorithm that is nearest to the proposed Plebeian Algorithm; however, there are subtle yet notable differences. The Plebeian Algorithm does not incorporate any notions of anonymity nor ephemerality. Content must be both traceable and permanently recorded. This will help assure that the goals of the social media companies at-large (which often differ from the goals of 4chan) remain consistent. Keeping the goals consistent for each individual social media platform will be essential to ensure that the users of the platform remain loyal while gaining the added benefits that the Plebeian Algorithm offers.

For 4chan's algorithm to become a Plebeian Algorithm, it should remove its ephemerality. This would be essential to ensure that content has the time to undergo the process. Content posted on 4chan's /b/ often lasts less than 1 minute [10]. As such, the Plebeian Algorithm would not have the time to undergo both two phases (the Flag and Jury Phases), a critical step that is necessary to the algorithm's democratic approach.

### Facebook

Facebook is a valuable selection, demonstrating a powerful social media platform and a tailored user experience; its popularity makes it useful for analysis. Although the Facebook company is not wholly transparent [9], the company has announced it highly favors personalized content of users (eg, posts from close friends and private groups) to that of public groups and pages to which the user likes and follows [33,34]. This presented a limitation for the analysis of Facebook data, as ideally, data had to be collected through preselected data sets for quantitative analysis [24].

There are many differences between the Facebook algorithm and the Plebeian Algorithm. The method used by Facebook, particularly during the COVID-19 pandemic, aims to combat the spread of misinformation and is based on neural networks trained to search for key terms in textual elements (including posts, comments, and statuses). It should also be noted that Facebook's algorithm includes a large amount of human work, which is easily biased. As has been stated in prior sections, there are several issues with this method of misinformation censorship; most notably, the Facebook algorithm is limited in scope to a specific subset of misinformation topics. Algorithms of this nature will detect specific key terms such as "COVID" included in the text of a post to provide additional information and resources for viewers; these algorithms will also provide information to the user sharing the post before publicizing the post. On the contrary, the Plebeian Algorithm is proactive in nature: it is universally applicable to all forms of misinformation and works to combat infodemics before they become widespread. As stated previously, infodemics of misinformation have led to the problem of pandemics becoming exacerbated and thus harder to control for public health workers. By implementing the Plebeian Algorithm, public health will be improved, especially concerning future pandemics, as potentially dangerous misrepresentations or falsehoods about the situation



XSL•FO
RenderX

will be contained to a smaller percentage of the populace, and thus ensures that reliable and trustworthy information is more accessible and widespread. The Plebeian Algorithm also requires less maintenance by developers, actively running automatically without the requirement of hard coding key terms to flag.

For Facebook to implement a Plebeian Algorithm, a high degree of planning would be required. Since Facebook is the most prevalent social media platform, a gradual implementation based on a rolling basis is recommended. AB testing should be used to ensure a smooth and successful implementation. Facebook should automate and democratize their home page algorithm to implement a Plebeian Algorithm for its service.

### *YouTube*

Due to the inherent difficulty in performing visual sentiment analysis for videos, comments of YouTube videos were analyzed. This does not give a complete picture of the YouTube algorithm, which attempts to keep users engaged longer on the website by presenting a tailored feed; the end goal being that the algorithm can predict videos the user would like to watch before they search [35]. This algorithm looks at a range of user data including watch time, closing a video tab, the user's interests, freshness, and user interactions with the video [35]. This algorithm has proven highly effective at finding and distributing viral content.

By aligning with the user's sentiment, the algorithm can effectively produce more positive comments, as seen in Figure 8. A sentiment filter used by YouTube includes the removal of videos that do not meet advertiser guidelines [36]. The main difference between YouTube's current content moderation approach and a Plebeian Algorithm's implementation of content moderation is the democratic aspect of content removal. This is made clear with many of YouTube's controversies within their community revolving around a lack of communication and censorship of larger creators [37].

The moderation system of YouTube is already a form of the Plebeian Algorithm with users being able to like and dislike comments or videos, in addition to reporting them if they are unwanted. The main disconnect between this and the Plebeian Algorithm is that, when a comment or video is reported, there is no public jury phase where the community decides if it stays. This has become clear with YouTube's controversies within the community revolving around issues such as the lack of communication and censorship of YouTube influencer Logan Paul. Should YouTube implement the Plebeian Algorithm, a Jury Phase is required after content is reported and prior to its removal process. It should also be noted that YouTube's jurors are not a random distribution. The moderation algorithms are programmed by humans, and as such, it is extremely difficult to ensure that the correct decisions are consistently being made. Artificial intelligence forms the basis of the YouTube algorithm, but the plebeian jury is replaced with a judge, who may be easily persuaded or hold personal biases. The wisdom of the crowd phenomenon (quod vide) plays a substantial role in the use of the jury for the Plebeian Algorithm.

### *Parler*

Parler uses a more typical algorithm. It limits posts to 1000 characters and circulates them to the user base at-large. Thus, unlike Reddit and 4chan, there are no communities in which content is posted on Parler. Parler was founded as a promoter of the freedom of speech, and as such, its user base is highly concerned with a lack of censorship on their posts [11].

Although this may at face value appear to be in direct opposition to the implementation of any algorithm, it is important to note that the Plebeian Algorithm ensures that any and all decisions regarding the verity of information remain in the users' hands. Parler would still benefit from implementing a Plebeian Algorithm, as it would preserve Parler's ultimate goal (to promote freedom of expression) while limiting the spread of misinformation.

For Parler to implement a Plebeian Algorithm, it must implement both the Flag and Jury Phases of the Plebeian Algorithm. Notably, the preservation of the freedom of expression on the platform must be ensured above all. This will ensure that the user base remains loyal and supportive of the change and does not boycott Parler or switch to a new social media platform (as they have already migrated from Twitter). The Parler user base is notably precarious, and it must ensure that the user base remains loyal to the platform. This should be done through proper marketing of the transition, which is to be discussed later.

### *Twitter*

Finally, Twitter uses a similar algorithm (in opposition still to Reddit and 4chan) in that posts and content are released to the user base at-large. The techniques of this algorithm particularly means that misinformation is more likely to spread on Twitter (and Parler) compared to other platforms (eg, Reddit and 4chan). The large user base on Twitter and the widespread availability of data must be taken into account, as it will be crucial that the culture and atmosphere of Twitter are maintained to ensure that the user base remains content with any algorithmic changes. Twitter's executives most likely would be interested in increasing their reach by attempting to regain the trust of those who migrated to Parler. These individuals are highly concerned with a decrease in censorship and an increase in the freedom of expression. They believe that the social media platform should remain separate from the process of promotion and demotion of content [11].

For Twitter to implement a Plebeian Algorithm, it must attempt to promote the freedom of expression and a decrease in censorship while also maintaining their reliability. This is done using the Plebeian Algorithm, which takes advantage of both concerns. Layperson algorithms have proven effective at curbing the spread of misinformation and at increasing reliability [18]. According to the definition by Epstein et al [16], the Plebeian Algorithm would be classified as a type of layperson algorithm. Twitter would need to place a level of trust in the layperson to provide the user base with liberty while maintaining truth in the content posted.



XSL•FO
RenderX

### Condorcet's Jury Theorem

As was researched in the 18th century by the Marquis de Condorcet, the Condorcet's jury theorem [38] clearly justifies the need for the Jury Phase in the Plebeian Algorithm. The theorem describes the behavior of a larger number of individuals selected to sit on a jury to judge the crimes of another individual. Proposed by Condorcet (and later proven by numerous mathematicians and statisticians in the late 20th century), the theorem explains that two scenarios may unfold when attempting to determine the truth by means of polling a sample of the population [38]. First, if the sample's understanding of the topic is poor, their judgement will not be certain. In this situation, the optimal sample size would be a single individual, as increasing the number of jurors will only increase the uncertainty [38].

However, a jury implemented in the Plebeian Algorithm should take Condorcet's jury theorem into account, by ensuring that the jury falls into the second scenario. The second type of jury would occur when the jury's knowledge of the subject is relatively high or is perceived as relatively high [38]. As such, an optimal Plebeian model would be passive, instead of aggressive, in its UI/UX. It should be ensured that acting as a juror is entirely optional and is opt-in instead of opt-out. The user interface should be minimal to ensure that the public reception of the implementation of the Plebeian Algorithm is positive. Although this will likely decrease the percentage of the sample who opt-in to act as jurors, consistence will be achieved due to the positive reception of the algorithm implementation. An ideal Plebeian Algorithm implementation to secure the second subset of juried defined by Condorcet's jury theorem may go unnoticed for the average user.

Assuming that an implementation of the Plebeian Algorithm can secure its jury into the latter jury type, it would secure the wisdom of the crowd. Increasing the sample selected as potential jurors will increase the certainty. This phenomenon has been described as "wisdom of the crowd" [39]. As the sample size increases, the certainty of the decision that the jury comes to also increases. Thus, taking a sample size of 1% has a higher possibility of accidentally selecting a group of the most extreme individuals, compared to randomly selecting a sample size of 10%.

### Eradication Versus Containment

One of the benefits of the Plebeian Algorithm in comparison to the status quo is the difference between eradication and containment. The algorithms of the current system tend to use an eradication approach. They view the issue of the spread of misinformation with a narrow perspective [35], and as such, they tend to implement a "Point-And-Shoot" algorithm. With this system, media companies determine which posts contain misinformation and eradicate them on a case-by-case basis. For example, many social media companies use a COVID-19 key term search and flag posts that contain them. They then link to a government website with information on the pandemic to the post.

This is in stark contrast to the Plebeian Algorithm, which takes a containment-based approach to the spread of misinformation. It should be noted that the technology to remove all instances of misinformation does not exist [40]. Instead, it is important that the algorithm detects as many cases of misinformation as possible and brings the rest to the broad public. This essentially "pops" any filter bubbles and echo chambers [26]. It allows for positive discussion from the community, which tends to lead toward a decrease in misinformation [4].

### Reduced Censorship

Another massive benefit of the Plebeian Algorithm is the reduction of censorship. With regard to the COVID-19 pandemic, a majority of misinformed posts has been spread by those with politically right ideologies, or Republicans. A total 48% of US Republicans believe that SARS-CoV-2 is no more dangerous than the common influenza [41] (compared to 25% of US Democrats [41]), and 42% of Republicans believe that hydroxychloroquine—a treatment for malaria—is an effective treatment for SARS-CoV-2 [41] (compared to 5% of Democrats [41]). Furthermore, Republicans (or those with political right ideologies) tend to be more concerned with the preservation of the freedoms of speech and expression. Thus, it is evident that we must preserve these freedoms for any algorithmic change to be effective. The Plebeian Algorithm goes further than this: it works to increase the rights of the individuals with respect to cross-region community matching freedom of expression. Individuals have the right to post and speak as they please and promote the spread of the information they deem to be pertinent. Additionally, they have the right to decide what content they want to see on the platform and what they do not. These benefits will help to ensure that the public reacts in a positive light to the change. Implementing a Plebeian Algorithm is a net positive; it is a positive change for both the containment of infodemics and the promotion of freedom of expression among social media users. Furthermore, under the Plebeian Algorithm, social media companies are still permitted to analyze user activity according to their privacy policies to provide appropriate advertisements tailored to the user. This will ensure that the revenues for social media companies will not be reduced in the process.

### Viral Naming Conventions

One subtopic explored herein is viral naming conventions and the connection between the name used to describe COVID-19 in relation to the level of verity in social media posts. For the purpose of this analysis, only posts on COVID-19 were considered; thus, social media platforms for which the data used herein was collected before 2020 were not analyzed (ie, Facebook and YouTube). Parler was examined at length since it uses a relatively standard social media algorithm, comparable to that of Twitter. A pickle file of Parler data, filtered to COVID-19, was generated from the Parler data set [9]. Additional filters were applied to the pickle file, as are described later.

To simplify the analysis, three categories of parleys were created on which analysis was performed separately. First, all posts mentioning COVID-19 using any naming conventions were gathered. The posts were collected using no additional filters, labelled "None." Second, from the COVID-19 parleys, a filter was applied to gather all parleys containing viral names referring to locations including, but not limited to, "Wuhan Virus," "China Virus," and "Indian Variant." All these terms have been

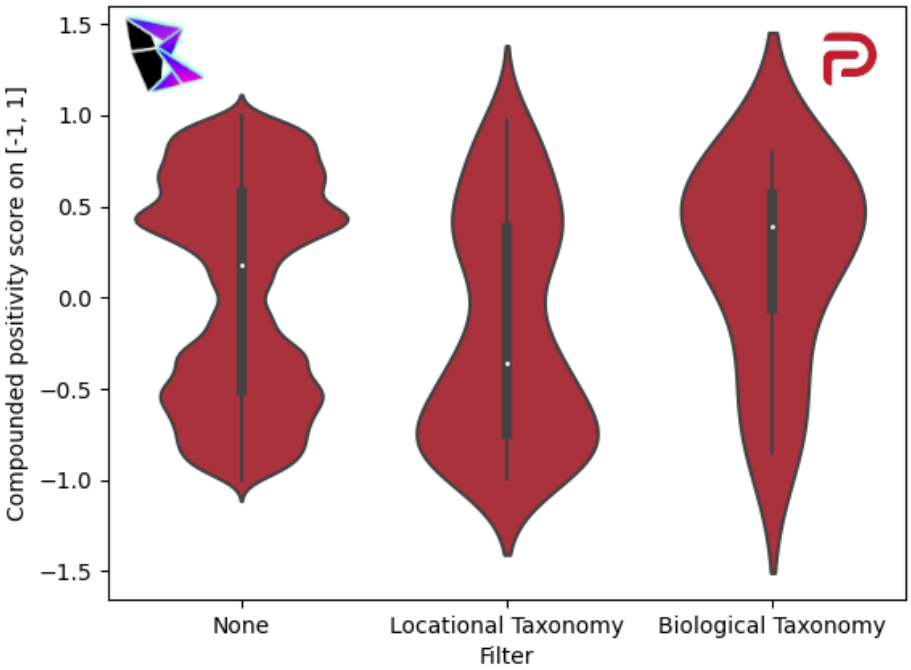


XSL•FO
RenderX

described by the US Centers for Disease Control and Prevention (CDC) to potentially propagate misinformation and xenophobia [42,43]. This filter was termed "Locational Taxonomy" [44]. The final filter, "Biological Taxonomy" [44], refers to the biological names for COVID-19, or officially approved names by the WHO, including, but not limited to, "SARS-CoV-2," "Alpha Variant," and "B.1.617." This nomenclature is used and promoted by the CDC [38] to limit xenophobia. As such, it was hypothesized that parleys using these terms would be less likely to be misinformative and more likely to have a positive sentiment [45].

Sentiment analysis was performed on all three filtered data sets, and the results were plotted as a violin plot in Figure 13. Each subplot portrays the three filters as discrete categories along the x-axis (ie, "None," "Locational Taxonomy," and "Biological Taxonomy") against the compounded polarity score on [–1, 1]. Each subplot visualizes a KDE that is rotated vertically for ease of visualization and a pictorial representation of the median, mean, first quartile, and third quartile.

**Figure 13.** Viral naming conventions on Parler: violin plot.




This visualization provides exceptionally relevant results. The data filtered to COVID-19 at-large was similar to the KDEs plotted for all of the social media algorithms discussed in the Results section, demonstrating a precise bimodal distribution with a positive and negative peak, and a neutral trough. The violin plots for the locational and biological taxonomies verified the hypothesis. The locational taxonomy filter showed strong negative sentiment, implying a higher likelihood of misinformation. In contrast, the biological taxonomy filter showed a strong positive sentiment, implying a greater degree of verity.

It is important to note that the findings are not limited to COVID-19. Similar findings were discovered (not pertaining to social media) relating to the 2009 H1N1/09 pandemic [46,47] and the 1918 Spanish Flu pandemic [15], among others [48,49], for which there were concerns surrounding xenophobic viral nomenclature. It is also pertinent to discuss the specific limitation surrounding the use of COVID-19 data for this analysis. It is often difficult for populations to alter their vocabulary to change the reference of a xenophobic initial name to an accurate descriptor [50]. Specifically, with regard to COVID-19 variants of concern (VoCs), many scientific sources still note the location of the VoC's discovery. This brings two significant points to the forefront of discussion: first, national health agencies need to provide precise and nonxenophobic nomenclature from the onset of pandemics, and second, the use of locational taxonomy should not automatically flag a post (ie, it should be flagged through sentimental analysis solely). The Plebeian Algorithm assists in this analysis, as it does not consider specific search terms, but rather pure sentiment. This handles issues surrounding truthful posts containing locational taxonomy. The lack of consensus among the scientific community should be noted regarding the potential benefits and drawbacks of using locational taxonomy [15,45].

## Geolocation

There exists a critical connection between the virtual and physical worlds as it pertains to the spread of misinformation and various consequences therefrom. Several studies have been conducted hereupon. One fundamental limitation posed by the use of social media platforms to track the spread of misinformation is the inability to deal with the spread of misinformation in more personal settings (eg, face-to-face interactions, videoconferencing, and direct messaging). Thus, a thorough study of the translation of misinformation from social media platforms to real-world phenomena will be conducted.

Myriad studies have been conducted surrounding infodemics [51-53]. This includes the correlation between geolocation of social media connections and various social determinants (eg, race, sex, and socioeconomic status) [51], and a study [54]



XSL•FO
RenderX

determining optimal methods of geolocation on social media. Two additional studies discussed the sociological consequences of geolocation in the context of social media, namely, the detection and reduction of youth cannabis consumption [52] and the applications of geolocation to urban planning [53].

Furthermore, studies suggest that there exists a strong correlation between trends on social media and events such as COVID-19 [12,51]. Evidently, any change on social media will have real-world impacts. Thus, it is apparent that a reduction in the amount of misinforming content in a social media user's home page corresponds with a reduction in the likelihood that they will propagate misinformative statements when having in-person conversations. Successful implementation of the Plebeian Algorithm will limit the spread of misinformation on social media platforms and in the lives of their users.

## Public Reaction

Skeptics of the Plebeian Algorithm might be concerned that such a massive alteration of the social media algorithm will incite hesitancy from the public. Whether this hesitance takes the form of negative feedback or boycotting, it is extremely legitimate and must be dealt with. Many will point to the 4chan platform as a negative example of an algorithm that offers user discretion regarding the promotion of content instead of a corporate algorithm.

### *Marketing*

This paper will first argue that the major difference between the two strategies lies within the realm of marketing. Marketing is a critical aspect of any social media company, especially when undergoing massive changes. In fact, some broad-scale social changes require marketing strategies [55]. Companies must ensure that the Plebeian Algorithm is adapted to meet the specific needs and goals of the social media company and its user base. For this reason, the Plebeian Algorithm is simply a suggested implementation, with a footnote that the algorithm must be highly adapted to the unique situation. Every social media company has varying objectives, such as Facebook's aim to connect friends, Reddit's goal to create conversations between like-minded individuals, and Parler's goal of preserving freedom of expression.

An effective marketing strategy for the transition to the Plebeian Algorithm ensures that users are aware that the overall atmosphere of the social media platform will not be altered. Promotion of the current atmosphere must take priority, lest the change face backlash by users. There is a potential, should improper marketing be implemented, that overly moderated individuals may leave the social media platform, leaving those with more extreme (and often misinformed) views to take over the widespread content of the platform. However, adequate marketing that emphasizes the static nature of the culture and social atmosphere of the platform during the transition alleviates this concern.

### *Feedback of Current Algorithms*

Second, this paper will discuss the feedback on the current algorithms as provided by the community. This feedback consists of discussions on social media platforms about each platform's algorithm. An analysis of preselected opinion pieces was performed [33,34,56-62]. These opinion pieces were sourced from well-known news or magazine sources, discussing the various social media platforms analyzed herein.

Overall, there is substantial desire for social media platforms to be more democratic in their algorithm. It is also widely believed among many social media users that, to improve algorithms, companies should implement a more transparent algorithm. Currently, algorithms vary widely and the functionality of most are not publicly available information. Changes improving transparency tend toward positive user feedback on the platform.

It is also critical to note that, for any implementation of the Plebeian Algorithm, a post must exceed a popularity threshold to be flagged in the Flag Phase. It is essential for the social media platforms to adapt their current algorithm to the determination of this popularity threshold. The goal of most current algorithms is to show users popular content that they may enjoy based on past interests. This can be done through a plethora of metrics, including likes, views, comments, recency of the post (termed "freshness") [35], and more. For example, the Twitter algorithm tends to prioritize the number of comments, whereas the YouTube algorithm prioritizes freshness.

## Implementation

Another concern of a potential implementing platform of the Plebeian Algorithm would be the technological requirements of the implementation, including storage and processing power required to conduct the Plebeian Algorithm on their millions of posts. Furthermore, this application of the Plebeian Algorithm would need to be a continuous process, ensuring that the algorithm continually updates when new comments are added to a post. As has been shown herein, the inclusion of comments increases the level of detail. All the data analyses visualized herein included comments and the text of the original post. As such, the computational power required appears to be great. There are, however, many alterations that can be made to the Plebeian Algorithm to reduce computation costs.

First, the Plebeian Algorithm does not need to be updated with the post of every new comment. It can be performed on intervals, whereby a subsection of posts is checked for new comments at every time interval. These new comments (and only the new comments) are then sent through sentiment analysis. In terms of data storage, it may prove useful for the social media platform to store a single additional byte of data for each post. The bit of highest significance, referred to as the "Flag of Need Determination," represented as $\varphi_d$ can be defined using the following equation:

such that:

where $n_d$ represents the value of determination (which is not scaled), $x$ represents a thread, $x_i$ represents a specific comment



XSL•FO
RenderX

or post within a thread, ν is a Boolean function returning a high value if the comment is new and low if it has been analyzed, sgn represents the signum function, $v_{act}$ and $v_{thres}$ represent the actual and threshold popularity of a thread in number of views, and $N$ represents the number of posts or comments in the thread.

If high, the post or thread can be safely skipped by the algorithm. If low, the post or thread will be analyzed to ensure that no misinformation goes undetected. The remaining seven bits of the data represent the sentiment of the entire thread, represented using $\beta_N$, where $N$ is the number of comments in the post, excluding the original post. These bits can be calculated using the following equations:

In some circumstances, it may be more computationally convenient to calculate $\beta_N$ recursively, which may be done using the following:

These equations demonstrate that a byte can be associated with each thread to decrease the processing requirements to execute the Plebeian Algorithm on a large scale.

It should also be noted that the Plebeian Algorithm is a machine learning model. It can be built to work in tandem with existing machine learning algorithms, thus decreasing the computing power required. Data storage is minimized using the one-byte storage method previously described. As is the case with all neural networks, the Plebeian Algorithm's Flag Phase will increase in accuracy over time by manipulating the string data as a validation set. Thus, the neural network will improve in accuracy over time. Due to time and resource limitations, the paper used the VADER; however, to increase Flag Phase precision over time, it is recommended that platforms implement the VADER sentiment analysis tool initially but build on it to adapt to the specific lexicon of the social media platform at the period of time. This accounts for minor differences in various social media algorithms and for lexical changes over time.

It is critical that a public release of the Plebeian Algorithm should be done through a process of AB testing. To efficiently fix any inevitable bugs present in the implementation of the algorithm (including any potential philosophical issues surrounding a specific realization/implementation), AB testing will be vital in the assurance that users consuming media under the new algorithm remain loyal to the brand and minimize any potential negative impacts. It will allow user feedback to be gathered for the small subsection of users presented with the Plebeian Algorithm implementation.

## Limitations

Although the Plebeian Algorithm is a great replacement for the current attempts by social media platforms to reduce the spread of misinformation, it is limited by several key factors. First, as stated earlier, the algorithm was only confirmed applicable for strictly text-based social media platforms and posts. Thus, the moderation of videos or images are outside of the scope of its use. Second, private sources of media such as chat rooms and servers are not within the scope of the algorithm, and thus, the algorithm is limited to public communication media. Third, the determination of a popularity threshold can be problematic. On Twitter, for example, a significant number of retweets are done passively (ie, they are not done for the express purpose of sharing with others but are done subconsciously by the user). Passive sharing may cause issues in the determination of whether a piece of content meets a popularity threshold. Finally, it is limited in the sense that it cannot determine what is misinformation at an instantaneous time selection, and as such, misinformation cannot be extracted from the algorithm at any time.

## Conclusions

The implications of this research are significant as to provide social media platforms with a new flagging method that uses sentiment analysis. This will be critical in the detection and prevention of infodemics and using a democratic approach that gives the power to the social media user to ultimately decide what content should be on the platform based on accuracy. The Plebeian Algorithm directly reduces political polarization and extremist ideas, which create a divide among users and improve cooperation on resolving key issues and problems plaguing humanity and restoring the trust between the public and experts.

Additionally, it is predicted that this will result in more reliable social media platforms, leading to an overall reduction of ignorance and misinformed opinions among users. Finally, the model created will lead to users expressing themselves without concern of the political viewpoint of the social media platform. Inherently, this also minimizes the impact of external biases, such as political climate, as those who vote will be completely random and anonymous.

Many areas of research remained unanalyzed. These topics include, but are not limited to:

- Conducting a study on the use of the Plebeian Algorithm on a selection of social media platforms and detecting the amount of misinformation over time after its implementation (ie, a real-world tested example) that would then be compared to current methods used, such as the aforementioned "Point-And-Shoot" Algorithm
- Creating a type of sentiment analysis for graphical content that could examine the emotion within an image to determine if it could be misinformation (eg, Snapchat, Instagram, and TikTok) [63,64]
- Determining the spread of misinformation correlated with the spread of viruses—this could be useful in predetermining locations (and users by extension) who are at higher risk of being exposed to or expounding misinformation
- Exploring the applicability of the Plebeian Algorithm in surveillance contexts, including for criminal investigations, employee onboarding, and health care [65-68]
- Analyzing the spread of misinformation through online vendors such as Amazon or eBay. In particular, recent audits of Amazon (as of 2021) show a dangerous disregard for reliable information, for example, presenting vaccine misinformation books along with well-cited vaccine



XSL•FO
RenderX

information books in generic searches for vaccine information [69-73]
- Applying models of higher sophistication for data analysis and visualization (which requires access to more in-depth data), including term frequency–inverse document measures [74] and Levenshtein distances [75] among others [76]
- Examining the optimal method of implementation and integration for the Plebeian Algorithm with various existing networking systems and infrastructures
- Continuing analysis of data collected to corroborate to prior studies on behavioral impacts of the sentiment of informative posts on social media
- Analyzing the role of corporate social media platforms (ie, Slack) in the dissemination of misinformation, especially in private chat channels
- Examining the misinformation containment models using juries, including the jury system implemented by Wikipedia
- Analyzing the rise of audio-form content, including podcasts, Clubhouse, and Spotify Greenroom audio-chat rooms, for the potential spread of misinformation—many of these media are becoming increasingly influential sources of news and information for many [77]
- Exploring the connection between location-based social media apps (eg, Foursquare) at the spread of geographic misinformation [78]

COVID-19 has had substantial impacts upon modern society. Optimists hoped these impacts would prove to unite a polarized world in the spirit of cooperation and global security. Although this has happened, their hopeful unity to the political schism has not. The Plebeian Algorithm is not a vaccine for an infodemic; however, it is a treatment to help curb and prevent the virus of misinformation from continuing to spread and grow out of control. This has the critical side-effect of putting power back in the hands of the people and removing the potential domination of a single entity (eg, a social media company) who may be swayed by external forces when deciding if content should be removed. All in all, it is recommended that social media executives consider the implementation of a variation of the Plebeian Algorithm, explicitly modified to adapt to the specifics of the platform. This will help curb misinformation both with regard to the COVID-19 infodemic and to prevent future infodemics.

## Acknowledgments

The authors would like to acknowledge the assistance in ideation from Anish R Verma from the STEM Fellowship. In addition, the authors are grateful for the assistance provided by Vinayak Nair concerning refinement of arguments from a data science perspective. Sponsors for the Undergraduate Big Data Challenge 2021, including JMIR Publications, Roche, SAS, Canadian Science Publishing, Digital Science, and Overleaf, made it possible for this research to be conducted.

## Conflicts of Interest

None declared.

## References

1. Karim F, Oyewande AA, Abdalla LF, Chaudhry Ehsanullah R, Khan S. Social media use and its connection to mental health: a systematic review. Cureus 2020 Jun 15;12(6):e8627 [FREE Full text] [doi: 10.7759/cureus.8627] [Medline: 32685296]
2. Kim HH. The impact of online social networking on adolescent psychological well-being (WB): a population-level analysis of Korean school-aged children. Int J Adolescence Youth 2016 Jun 16;22(3):364-376 [FREE Full text] [doi: 10.1080/02673843.2016.1197135]
3. Watson A. Share of adults who use social media as a source of news in selected countries worldwide as of February 2020. Statista. 2020. URL: https://www.statista.com/statistics/718019/social-media-news-source/ [accessed 2021-07-14]
4. Walsh DR. Neutral isn't neutral: an analysis of misinformation and sentiment in the wake of the capitol riots. The Research Repository @ WVU. 2021 May 10. URL: https://researchrepository.wvu.edu/etd/8055/ [accessed 2021-07-14]
5. Wardle C, Derakhshan H. Information disorder: toward an interdisciplinary framework for research and policy making. Counc Europe Rep 2017 Dec 27;2017(09):1-109 [FREE Full text]
6. Bradd S. Infodemic. World Health Organization. URL: https://www.who.int/health-topics/infodemic#tab=tab_1 [accessed 2021-07-14]
7. Hazelwood K, Bird S, Brooks D, Chintala S, Diril U, Dzhulgakov D, et al. Applied machine learning at Facebook: a datacenter infrastructure perspective. Facebook Inc. 2017. URL: https://research.fb.com/wp-content/uploads/2017/12/hpca-2018-facebook.pdf [accessed 2021-07-14]
8. Berstein M, Monroy-Hernández A, Harry D, André P, Panovich K, Vargas G. 4chan and /b/: an analysis of anonymity and ephemerality in a large online community. 2011 Jul 17 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2021; Barcelona, Catalonia, Spain p. A URL: https://www.researchgate.net/publication/221297869_4chan_and_b_An_Analysis_of_Anonymity_and_Ephemerality_in_a_Large_Online_Community
9. Aliapoulios M, Bevensee E, Blackburn J, Bradlyn B, De Cristofaro E, Stringhini G, et al. An early look at the Parler online social network. arXiv. Preprint posted online on January 11, 2021 [FREE Full text]



XSL•FO
RenderX

10. Massanari A. #Gamergate and The Fappening: how Reddit's algorithm, governance, and culture support toxic technocultures. New Media Soc 2016 Jul 09;19(3):329-346. [doi: 10.1177/1461444815608807]
11. Kumar KPK, Geethakumari G. Detecting misinformation in online social networks using cognitive psychology. Hum-centric Computing Inf Sci 2014 Sep 24;4(1):1-22. [doi: 10.1186/s13673-014-0014-x]
12. Stephens M. A geospatial infodemic: mapping Twitter conspiracy theories of COVID-19. Dialogues Hum Geography 2020 Jun 23;10(2):276-281. [doi: 10.1177/2043820620935683]
13. Fernández M, Bellogín A, Cantadlr I. Analysing the effect of recommendation algorithms on the amplification of misinformation. arXiv. Preprint posted online on March 26, 2021 [FREE Full text]
14. Plaza M, Paladino L, Opara I, Firstenberg M, Wilson B, Papadimos T, et al. The use of distributed consensus algorithms to curtail the spread of medical misinformation. Int J Academic Med 2019;5(2):93. [doi: 10.4103/ijam.ijam_47_19]
15. Hoppe T. "Spanish Flu": when infectious disease names blur origins and stigmatize those infected. Am J Public Health 2018 Nov;108(11):1462-1464. [doi: 10.2105/AJPH.2018.304645] [Medline: 30252513]
16. Epstein Z, Pennycook G, Rand D. Will the crowd game the algorithm? Using layperson judgments to combat misinformation on social media by downranking distrusted sources. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Apr 25 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI URL: https://dl.acm.org/doi/10.1145/3313831.3376232 [doi: 10.1145/3313831.3376232]
17. Kaushik S. Covid vaccine tweets. Kaggle. 2021. URL: https://www.kaggle.com/kaushiksuresh147/covidvaccine-tweets [accessed 2021-07-14]
18. Memon SA, Carley KM. CMU-MisCov19: a novel Twitter dataset for characterizing COVID-19 misinformation. Zenodo. 2020 Sep 19. URL: https://zenodo.org/record/4024154 [accessed 2021-07-14]
19. Gruzd A, Mai P. Inoculating against an infodemic: a canada-wide Covid-19 news, social media, and misinformation survey. SSRN J. Preprint posted online on May 11, 2020 . [doi: 10.2139/ssrn.3597462]
20. Palachy S. Twitter datasets. GitHub. 2020. URL: https://github.com/shaypal5/awesome-twitter-data [accessed 2021-07-14]
21. Tankovska H. Number of monthly active Facebook users worldwide as of 1st Quarter 2021. Statista. 2021. URL: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook\-users-worldwide [accessed 2021-07-14]
22. Span J. FacebookR comments. GitHub. 2017. URL: https://github.com/jerryspan/FacebookR [accessed 2021-07-14]
23. Mitchell J. Trending YouTube video statistics and comments. Kaggle. 2017. URL: https://www.kaggle.com/datasnaek/youtube [accessed 2021-07-14]
24. Rhodes S. Filter bubbles, echo chambers, and fake news: how social media conditions individuals to be less critical of political misinformation. Political Commun 2021 May 01:1-22 [FREE Full text] [doi: 10.1080/10584609.2021.1910887]
25. Wollebæk D, Karlsen R, Steen-Johnsen K, Enjolras B. Anger, fear, and echo chambers: the emotional basis for online behavior. Soc Media Soc 2019 Apr 09;5(2):205630511982985. [doi: 10.1177/2056305119829859]
26. Pariser E. The filter bubble: what the internet is hiding from you. Standford HCI Group. 2011. URL: https://hci.stanford.edu/courses/cs047n/readings/The_Filter_Bubble.pdf [accessed 2021-07-14]
27. Frerichs RR. Simple random sampling. Jonathan and Karin Fielding School of Public Health. URL: https://www.ph.ucla.edu/epi/rapidsurveys/RScourse/RSbook_ch3.pdf [accessed 2021-07-14]
28. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. CJEM 2020 Jul 06;22(4):418-421 [FREE Full text] [doi: 10.1017/cem.2020.361] [Medline: 32248871]
29. Auerbach D. Reddit Scandals: Does Reddit Have a Transparency Problem? Slate. URL: https://slate.com/technology/2014/10/reddit-scandals-does-the-site-have-a-transparency-problem.html [accessed 2021-07-14]
30. DeVito M, Gergle D, Bernholtz J. Algorithms ruin everything: RIPTwitter, folk theories, and resistance to algorithmic change in social media HCI and collective action. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 May 02 Presented at: CHI '17; May 6-11, 2017; Denver, CO p. 3163-3174. [doi: 10.1145/3025453.3025659]
31. Comparative study of HITS and PageRank Link based ranking algorithms. Cornell University Blog Service. 2015 Oct 27. URL: https://blogs.cornell.edu/info2040/2015/10/27/comparative-study-of-hits-and-pagerank-link-based-ranking-algorithms/ [accessed 2021-07-14]
32. McIntyre MM. Relational agency, networked technology, and the social media aftermath of the Boston Marathon bombing. ProQuest. 2015 Sep 16. URL: https://www.proquest.com/openview/f3b73d2fae6805d92eed81229392053b/1?cbl=18750&parentSessionId=hkM46Px9ikwBr%2FfL65fe%2FaUV3N0iqN0Q4j6bazFyTvE%3D&pq-origsite=gscholar&accountid=47192 [accessed 2021-07-14]
33. Hutchinson A. The pros and cons of Facebook's coming news feed changes-from a page perspective. Social Media Today. 2018 Jan 17. URL: https://www.socialmediatoday.com/news/the-pros-and-cons-of-facebooks-coming-newsfeed-changes-from-a-page-pers/514776/ [accessed 2021-07-14]
34. Cooper P. How the Facebook algorithm works in 2021 and how to make it work for you. Hootsuite Blog. 2021 Feb 10. URL: https://blog.hootsuite.com/facebook-algorithm/ [accessed 2021-07-14]
35. Covington P, Adams J, Sargin E. Deep neural networks for YouTube recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems. 2016 Sep 15 Presented at: RecSys '16; September 15-19, 2016; Boston, MA p. 1-8




XSL•FO
RenderX

URL: https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45530.pdf [doi: 10.1145/2959100.2959190]
36. Terms of service. YouTube. 2021. URL: https://www.youtube.com/static?template=terms [accessed 2021-07-14]
37. Southerton C, Marshall D, Aggleton P, Rasmussen ML, Cover R. Restricted modes: Social media, content classification and LGBTQ sexual citizenship. N Media Soc 2020 Feb 13;23(5):920-938. [doi: 10.1177/1461444820904362]
38. Austen-Smith D, Banks JS. Information aggregation, rationality, and the Condorcet jury theorem. Am Polit Sci Rev 2014 Aug 01;90(1):34-45. [doi: 10.2307/2082796]
39. Yi S, Steyvers M, Lee MD, Dry MJ. The wisdom of the crowd in combinatorial problems. Cogn Sci 2012 Apr;36(3):452-470. [doi: 10.1111/j.1551-6709.2011.01223.x] [Medline: 22268680]
40. Ghoshal AK, Das N, Das S. Influence of community structure on misinformation containment in online social networks. Knowledge-Based Syst 2021 Feb;213:106693. [doi: 10.1016/j.knosys.2020.106693]
41. Cox DA, Halpin J. Conspiracy theories, misinformation, COVID-19, and the 2020 election tech. The Survey Center on American Life. 2020 Oct 13. URL: https://www.americansurveycenter.org/research/conspiracy-theories-misinformation-covid-19-and-the-2020-election/ [accessed 2021-07-14]
42. Abdool Karim SS, de Oliveira T, Loots G. Appropriate names for COVID-19 variants. Science 2021 Mar 19;371(6535):1215. [doi: 10.1126/science.abh0836] [Medline: 33737481]
43. Scientific Nomenclature. Centers for Disease Control and Prevention. 2014. URL: https://wwwnc.cdc.gov/eid/page/scientific-nomenclature [accessed 2021-07-14]
44. Hull R, Rima B. Virus taxonomy and classification: naming of virus species. Arch Virol 2020 Nov;165(11):2733-2736. [doi: 10.1007/s00705-020-04748-7] [Medline: 32740831]
45. Masters-Waage TC, Jha N, Reb J. COVID-19, coronavirus, Wuhan virus, or China virus? Understanding how to "Do No Harm" when naming an infectious disease. Front Psychol 2020;11:561270. [doi: 10.3389/fpsyg.2020.561270] [Medline: 33362626]
46. Vigsø O. Naming is framing: swine flu, new flu, and A(H1N1). Observatorio 2010 Jan;4:1.
47. McCauley M, Minsky S, Viswanath K. The H1N1 pandemic: media frames, stigmatization and coping. BMC Public Health 2013 Dec 03;13:1116 [FREE Full text] [doi: 10.1186/1471-2458-13-1116] [Medline: 24299568]
48. Sell TK, Hosangadi D, Trotochaud M. Misinformation and the US Ebola communication crisis: analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak. BMC Public Health 2020 May 07;20(1):550 [FREE Full text] [doi: 10.1186/s12889-020-08697-3] [Medline: 32375715]
49. Singh RP, Valkonen JPT, Gray SM, Boonham N, Jones RAC, Kerlan C, et al. Discussion paper: the naming of Potato virus Y strains infecting potato. Arch Virol 2008;153(1):1-13. [doi: 10.1007/s00705-007-1059-1] [Medline: 17943395]
50. Mallapaty S. Should virus-naming rules change during a pandemic? The question divides virologists. Nature 2020 Aug;584(7819):19-20. [doi: 10.1038/d41586-020-02243-2] [Medline: 32733098]
51. Baucom E, Sanjari A, Liu X, Chen M. Mirroring the real world in social media: Twitter, geolocation, and sentiment analysis. In: Proceedings of the 2013 International Workshop on Mining Unstructured Big Data Using Natural Language Processing. 2013 Oct Presented at: UnstructureNLP '13; October 28, 2013; San Francisco, CA p. 61-68 URL: https://dl.acm.org/doi/10.1145/2513549.2513559 [doi: 10.1145/2513549.2513559]
52. Chung T, Pelechrinis K, Faloutsos M, Hylek L, Suffoletto B, Feldstein Ewing SW. Innovative routes for enhancing adolescent marijuana treatment: interplay of peer influence across social media and geolocation. Curr Addict Rep 2016 Apr 1;3(2):221-229. [doi: 10.1007/s40429-016-0095-x]
53. Milusheva S, Marty R, Bedoya G, Williams S, Resor E, Legovini A. Applying machine learning and geolocation techniques to social media data (Twitter) to develop a resource for urban planning. PLoS One 2021;16(2):e0244317 [FREE Full text] [doi: 10.1371/journal.pone.0244317] [Medline: 33534801]
54. Williams E, Gray J, Dixon B. Improving geolocation of social media posts. Pervasive Mobile Computing 2017 Apr;36:68-79. [doi: 10.1016/j.pmcj.2016.09.015]
55. Kotler P, Zaltman G. Social marketing: an approach to planned social change. J Mark 1971 Jul;35(3):3-12. [Medline: 12276120]
56. Mayfield D. Social media algorithms 2021: updates and tips by platform. StoryChief. URL: https://storychief.io/blog/social-media-algorithms-updates-tips [accessed 2021-07-14]
57. Settlage B. Pros and cons of the social media algorithm age. Amplimark. 2021 Feb 17. URL: https://www.amplimark.com/pros-and-cons-of-the-social-media-algorithm-age/2021 [accessed 2021-07-14]
58. Disadvantages of the Facebook algorithm and how you can workaround it. On The Maps! Digital Marketing Company. 2020 Jul 01. URL: https://onthemaps.com/disadvantages-of-the-facebook-algorithm-and-how-you-can-workaround-it/ [accessed 2021-07-14]
59. Thottam I. Pros and cons of Twitter's new algorithmic timeline. Paste Magazine. 2016 Feb 17. URL: https://www.pastemagazine.com/tech/twitter/pros-and-cons-of-twitters-new-algorithmic-timeline/ [accessed 2021-07-14]
60. Curvelo R. The pros and Ccons of using Twitter as part of your digital marketing strategy. Matrix Internet. 2020 Feb 28. URL: https://www.matrixinternet.ie/the-pros-and-cons-of-twitter/ [accessed 2021-07-14]



XSL•FO
RenderX

61. Casper H. Everything you need to know about Twitter's timeline algorithm. Click to Tweet. 2017 Mar 14. URL: https://clicktotweet.com/blog/everything-you-need-to-know-about-twitter-timeline-algorithm [accessed 2021-07-14]
62. Beasing D. Does Facebook's new algorithm add up for radio? Radio World. 2018 Mar 02. URL: https://www.radioworld.com/news-and-business/does-facebooks-new-algorithm-add-up-for-radio [accessed 2021-07-14]
63. Agung N, Darma G. Opportunities and challenges of Instagram algorithm in improving competitive advantage. Int J Innovative Sci Res Technol 2019 Jan;4(1):1.
64. Alumnus Marc Faddoul discovers racial biases in TikTok's algorithm. UC Berkeley School of Information. URL: https://www.ischool.berkeley.edu/news/2020/alumnus-marc-faddoul-discovers-racial-biases-tiktoks-algorithm [accessed 2021-07-14]
65. Mateescu A, Brunton D, Rosenblat A, Patton D, Gold Z, Boyd D. Social media surveillance and law enforcement. Data Civil Rights N Era Policing Justice 2015 Oct 27:1-11 [FREE Full text]
66. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine learning: a systematic review. J Biomed Inform 2020 Aug;108:103500 [FREE Full text] [doi: 10.1016/j.jbi.2020.103500] [Medline: 32622833]
67. Bizzi L. Should HR managers allow employees to use social media at work? Behavioral and motivational outcomes of employee blogging. Int J Hum Resource Manage 2017 Nov 14;31(10):1285-1312. [doi: 10.1080/09585192.2017.1402359]
68. Alexander E, Mader D, Mader F. Using social media during the hiring process: a comparison between recruiters and job seekers. J Global Scholars Marketing Sci 2019 Jan 14;29(1):78-87 [FREE Full text] [doi: 10.1080/21639159.2018.1552530]
69. Shin J, Valente T. Algorithms and health misinformation: a case study of vaccine books on Amazon. J Health Commun 2020 May 03;25(5):394-401. [doi: 10.1080/10810730.2020.1776423] [Medline: 32536257]
70. Juneja P, Mitra T. Auditing e-commerce platforms for algorithmically curated vaccine misinformation. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 May 06 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan p. 1-27 URL: https://dl.acm.org/doi/10.1145/3411764.3445250 [doi: 10.1145/3411764.3445250]
71. Smith B, Linden G. Two decades of recommender systems at Amazon.com. IEEE Internet Computing 2017 May;21(3):12-18. [doi: 10.1109/mic.2017.72]
72. Krishnamurthy S. A Comparative Analysis of eBay and Amazon. Seattle, WA: University of Washington; 2004:29-44.
73. Yuan TT, Chen Z, Mathieson M. Predicting eBay listing conversion. In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. 2011 Jul Presented at: SIGIR '11; July 24-28, 2011; Beijing, China URL: http://www.bayimage.com/code/Sigir2011_docs_p1335.pdf [doi: 10.1145/2009916.2010188]
74. Aizawa A. An information-theoretic perspective of tf–idf measures. Inf Processing Manage 2003 Jan;39(1):45-65. [doi: 10.1016/s0306-4573(02)00021-3]
75. Andoni A, Onak K. Approximating edit distances in near-linear time. In: Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing. 2009 Jul 02 Presented at: STOC '09; May 31-June 2, 2009; Bethesda, MD p. 199. [doi: 10.1145/1536414.1536444]
76. Hong I, Rutherford A, Cebrian M. Social mobilization and polarization can create volatility in COVID-19 pandemic control. Appl Netw Sci 2021;6(1):11 [FREE Full text] [doi: 10.1007/s41109-021-00356-9] [Medline: 33614902]
77. Alang N. Clubhouse chat beating social media trolls: for better or worse, site offers conversation among the like-minded. Toronto Star. 2021 Mar. URL: https://www.pressreader.com/canada/toronto-star/20210306/281900185950631 [accessed 2021-07-14]
78. Zhao Y, Nie L, Wang X, Chua T. Personalized recommendations of locally interesting venues to tourists via cross-region community matching. ACM Trans Intelligent Syst Technol 2014 Oct;5(3):1-26. [doi: 10.1145/2532439]

## Abbreviations

**API:** application programming interface
**CDC:** Centers for Disease Control and Prevention
**KDE:** kernel density estimation
**VADER:** Valence Aware Dictionary and Sentiment Reasoner
**VoC:** variant of concern
**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 27.07.21; peer-reviewed by M Lotto, N Seeman; comments to author 25.08.21; revised version received 15.09.21; accepted 26.11.21; published 21.12.21.*

*Please cite as:*
*Fedoruk B, Nelson H, Frost R, Fucile Ladouceur K*
*The Plebeian Algorithm: A Democratic Approach to Censorship and Moderation*
*JMIR Form Res 2021;5(12):e32427*
*URL: https://formative.jmir.org/2021/12/e32427*
*doi:10.2196/32427*
*PMID:34854812*



XSL•FO
RenderX

©Benjamin Fedoruk, Harrison Nelson, Russell Frost, Kai Fucile Ladouceur. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.



XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org



XSL•FO
RenderX