

---

# JMIR Formative Research

---

Impact Factor (2023): 2.0  
Volume 6 (2022), Issue 9 ISSN 2561-326X Editor in Chief: Amaryllis Mavragani, PhDc

---

## Contents

### Original Papers

- Measuring Daily Activity Rhythms in Young Adults at Risk of Affective Instability Using Passively Collected Smartphone Data: Observational Study (e33890)  
Benny Ren, Cedric Xia, Philip Gehrman, Ian Barnett, Theodore Satterthwaite. . . . . 7
- Measurement of Heart Rate Using the Withings ScanWatch Device During Free-living Activities: Validation Study (e34280)  
Oonagh Giggins, Julie Doyle, Suzanne Smith, Daniel Crabtree, Matthew Fraser. . . . . 22
- Digital Health Screening in People With HIV in Uganda to Increase Alcohol Use Reporting: Qualitative Study on the Development and Testing of the Self-administered Digital Screener for Health (e35015)  
Nneka Emenyonu, Allen Kekibiina, Sarah Woolf-King, Catherine Kyampire, Robin Fatch, Carol Dawson-Rose, Winnie Muyindike, Judith Hahn.  
3 1
- Integrating Social Determinants of Health With Tobacco Treatment for Individuals With Opioid Use Disorder: Feasibility and Acceptability Study of Delivery Through Text Messaging (e36919)  
Hasmeena Kathuria, Divya Shankar, Vinson Cobb, Julia Newman, Katia Bulekova, Scott Wertz, Belinda Borrelli. . . . . 49
- The Impact of Smartphone Apps Designed to Reduce Food Waste on Improving Healthy Eating, Financial Expenses and Personal Food Waste: Crossover Pilot Intervention Trial Studying Students' User Experiences (e38520)  
Therese Mathisen, Frode Johansen. . . . . 64
- The Use of Smartphone Serious Gaming Apps in the Treatment of Substance Use Disorders: Observational Study on Feasibility and Acceptability (e34159)  
Thelma Schilt, Elvira Ruijter, Nikky Godeschalk, Marit van Haaster, Anna Goudriaan. . . . . 76
- Transdiagnostic Internet-Delivered Cognitive Behavioral Therapy for Symptoms of Postpartum Anxiety and Depression: Feasibility Randomized Controlled Trial (e37216)  
Victoria Suchan, Vanessa Peynenburg, David Thiessen, Marcie Nugent, Blake Dear, Nikolai Titov, Heather Hadjistavropoulos. . . . . 88
- A Method to Deliver Automated and Tailored Intervention Content: 24-month Clinical Trial (e38262)  
Hailey Miller, Corrine Voils, Kate Cronin, Elizabeth Jeanes, Jeffrey Hawley, Laura Porter, Rachel Adler, Whitney Sharp, Samantha Pabich, Kara Gavin, Megan Lewis, Heather Johnson, William Yancy Jr, Kristen Gray, Ryan Shaw. . . . . 105
- Design, Development, and Testing of BEST4Baby, an mHealth Technology to Support Exclusive Breastfeeding in India: Pilot Study (e32795)  
Tony Ma, Katie Chang, Amal Alyusuf, Elina Bajracharya, Yukiko Washio, Patricia Kelly, Roopa Bellad, Niranjana Mahantashetti, Umesh Charantimath, Vanessa Short, Parth Lalakia, Frances Jaeger, Shivaprasad Goudar, Richard Derman. . . . . 118

<b>An Intervention Offering Self-management Support Through mHealth and Health Coaching to Patients With Prostate Cancer: Interpretive Description of Patients' Experiences and Perspectives (e34471)</b>	
Louise Obro, Palle Osther, Jette Ammentorp, Gitte Pihl, Kasper Heiselberg, Peter Krogh, Charlotte Handberg. . . . .	131
<b>Understanding Gender Biases and Differences in Web-Based Reviews of Sanctioned Physicians Through a Machine Learning Approach: Mixed Methods Study (e34902)</b>	
Julia Barnett, Margrét Bjarnadóttir, David Anderson, Chong Chen. . . . .	143
<b>The Helpfulness of Web-Based Mental Health and Well-being Forums for Providing Peer Support for Young People: Cross-sectional Exploration (e36432)</b>	
Emily Banwell, Terry Hanley, Santiago De Ossorno Garcia, Charlotte Mindel, Thomas Kayll, Aaron Sefi. . . . .	158
<b>Quality of Late-Life Depression Information on the Internet: Website Evaluation Study (e36177)</b>	
Teaghan Pryor, Kristin Reynolds, Paige Kirby, Matthew Bernstein. . . . .	173
<b>Home-Based Electronic Cognitive Therapy in Patients With Alzheimer Disease: Feasibility Randomized Controlled Trial (e34450)</b>	
Anna Marin, Renée DeCaro, Kylie Schiloski, Ala'a Elshaar, Brigid Dwyer, Ana Vives-Rodriguez, Rocco Palumbo, Katherine Turk, Andrew Budson. . . . .	186
<b>Predicting Depression in Patients With Knee Osteoarthritis Using Machine Learning: Model Development and Validation Study (e36130)</b>	
Zuzanna Nowinka, M Alagha, Khadija Mahmoud, Gareth Jones. . . . .	202
<b>Examining the Delivery of a Tailored Chinese Mind-Body Exercise to Low-Income Community-Dwelling Older Latino Individuals for Healthy Aging: Feasibility and Acceptability Study (e40046)</b>	
Yan Du, Neela Patel, Arthur Hernandez, Maria Zamudio-Samano, Shiyu Li, Tianou Zhang, Roman Fernandez, Byeong Choi, William Land, Sarah Ullevig, Vanessa Estrada Coats, Jessh Moussavou, Deborah Parra-Medina, Zenong Yin. . . . .	216
<b>Motive-Oriented, Personalized, Internet-Based Interventions for Depression: Nonclinical Experimental Study (e37287)</b>	
Lara Bucker, Thomas Berger, Alina Bruhns, Stefan Westermann. . . . .	234
<b>Development of a Mobile App for Self-Care Against COVID-19 Using the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) Model: Methodological Study (e39718)</b>	
Hamid Saeidnia, Marcin Kozak, Marcel Ausloos, Claudiu Herteliu, Zahra Mohammadzadeh, Ali Ghorbi, Mehrdad Karajizadeh, Mohammad Hassanzadeh. . . . .	249
<b>Personalized Energy Expenditure Estimation: Visual Sensing Approach With Deep Learning (e33606)</b>	
Toby Perrett, Alessandro Masullo, Dima Damen, Tilo Burghardt, Ian Craddock, Majid Mirmehdi. . . . .	263
<b>Digital Global Recruitment for Women's Health Research: Cross-sectional Study (e39046)</b>	
Erika Rodriguez, Komal Peer, Victoria Fruh, Kaitlyn James, Anna Williams, Alexis de Figueiredo Veiga, Michael Winter, Amanda Shea, Ann Aschengrau, Kevin Lane, Shruthi Mahalingaiah. . . . .	273
<b>Agreement Between Clinically Measured Weight and Self-reported Weight Among Patients With Type 2 Diabetes Through an mHealth Lifestyle Coaching Program in Denmark: Secondary Analysis of a Randomized Controlled Trial (e40739)</b>	
Albi Imeraj, Thomas Olesen, Ditte Laursen, Jens Søndergaard, Carl Brandt. . . . .	284
<b>The Usability of a Smartphone-Based Fall Risk Assessment App for Adult Wheelchair Users: Observational Study (e32453)</b>	
Mikaela Frechette, Jason Fanning, Katherine Hsieh, Laura Rice, Jacob Sosnoff. . . . .	297

Acceptability and Impact of an Educational App (iCare) for Informal Carers Looking After People at Risk of Pressure Ulceration: Mixed Methods Pilot Study (e36517)  
 Eamonn McKeown, Caroline McGraw, Pru Holder, Jenny Shand, Shashivadan Hirani. . . . . 310

Waiting Lists for Psychotherapy and Provider Attitudes Toward Low-Intensity Treatments as Potential Interventions: Survey Study (e39787)  
 Allison Peipert, Anne Krendl, Lorenzo Lorenzo-Luaces. . . . . 325

Influence of 2 Digital Exercise Modules of a Multimodular System on Balance and Leg Strength Under Consideration of Use Adherence: Prospective Cohort Study (e36805)  
 Verena Venek, Christina Kranzinger, Sonja Jungreitmayr, Susanne Ring-Dimitriou, Hermann Schwameder, Thomas Stöggel. . . . . 338

A Web-Based Positive Psychology App for Patients With Bipolar Disorder: Development Study (e39476)  
 Bart Geerling, Saskia Kelders, Anja Stevens, Ralph Kupka, Ernst Bohlmeijer. . . . . 354

Resilience in Web-Based Mental Health Communities: Building a Resilience Dictionary With Semiautomatic Text Analysis (e39013)  
 Yong-Bin Kang, Anthony McCosker, Peter Kamstra, Jane Farmer. . . . . 372

Relative Effectiveness of Social Media, Dating Apps, and Information Search Sites in Promoting HIV Self-testing: Observational Cohort Study (e35648)  
 Chrysovalantis Stafylis, Gabriella Vavala, Qiao Wang, Bethany McLeman, Shea Lemley, Sean Young, Haiyi Xie, Abigail Matthews, Neal Oden, Leslie Revoredo, Dikla Shmueli-Blumberg, Emily Hichborn, Erin McKelle, Landhing Moran, Petra Jacobs, Lisa Marsch, Jeffrey Klausner. . . . . 3 8 9

Comparing Professional and Consumer Ratings of Mental Health Apps: Mixed Methods Study (e39813)  
 Georgie Hudson, Esther Negbenose, Martha Neary, Sonja Jansli, Stephen Schueller, Til Wykes, Sagar Jilka. . . . . 401

Using Social Media to Facilitate Communication About Women’s Testing: Tool Validation Study (e35035)  
 Tara Coffin, Deborah Bowen, Karen Lu, Elizabeth Swisher, Nadine Rayes, Barbara Norquist, Stephanie Blank, Douglas Levine, Jamie Bakkum-Gamez, Gini Fleming, Olufunmilayo I Olopade, Iris Romero, Alan D’Andrea, Denise Nebgen, Christine Peterson, Mark Munsell, Kathleen Gavin, Jamie Crase, Deborah Polinsky, Rebecca Lechner. . . . . 413

An Economic Impact Model for Estimating the Value to Health Systems of a Digital Intervention for Diabetes Primary Care: Development and Usefulness Study (e37745)  
 Brenton Powers, Amy Bucher. . . . . 429

Exploring Socioeconomic Status as a Global Determinant of COVID-19 Prevalence, Using Exploratory Data Analytic and Supervised Machine Learning Techniques: Algorithm Development and Validation Study (e35114)  
 Luke Winston, Michael McCann, George Onofrei. . . . . 440

Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques (e36118)  
 Muhammad Kabir, Maisha Islam, Anika Kabir, Adiba Haque, Md Rhaman. . . . . 458

Comparing Transactional eHealth Literacy of Individuals With Cancer and Surrogate Information Seekers: Mixed Methods Study (e36714)  
 Taylor Vasquez, Carma Bylund, Jordan Alpert, Julia Close, Tien Le, Merry Markham, Greenberry Taylor, Samantha Paige. . . . . 471

Usability Testing of the Kidney Score Platform to Enhance Communication About Kidney Disease in Primary Care Settings: Qualitative Think-Aloud Study (e40001)  
 Delphine Tuot, Susan Crowley, Lois Katz, Joseph Leung, Delly Alcantara-Cadillo, Christopher Ruser, Elizabeth Talbot-Montgomery, Joseph Vassalotti. . . . . 489

<p>Comparison Between QT and Corrected QT Interval Assessment by an Apple Watch With the AccurBeat Platform and by a 12-Lead Electrocardiogram With Manual Annotation: Prospective Observational Study (e41241)</p> <p>Sara Chokshi, Gulzhan Tologonova, Rose Calixte, Vandana Yadav, Naveed Razvi, Jason Lazar, Stan Kachnowski. . . . .</p>	501
<p>Thinking Aloud or Screaming Inside: Exploratory Study of Sentiment Around Work (e30113)</p> <p>Marzia Hoque Tania, Md Hossain, Nuzhat Jahanara, Ilya Andreev, David Clifton. . . . .</p>	514
<p>Telehealth-Supported Decision-making Psychiatric Care for Suicidal Ideation: Longitudinal Observational Study (e37746)</p> <p>Erin O'Callaghan, Nicole Mahrer, Heather Belanger, Scott Sullivan, Christine Lee, Carina Gupta, Mirène Winsberg. . . . .</p>	536
<p>Attention-Based Models for Classifying Small Data Sets Using Community-Engaged Research Protocols: Classification System Development and Validation Pilot Study (e32460)</p> <p>Brian Ferrell, Sarah Raskin, Emily Zimmerman, David Timberline, Bridget McInnes, Alex Krist. . . . .</p>	552
<p>"There's No Heroin Around Anymore. It's All Fentanyl." Adaptation of an Opioid Overdose Prevention Counseling Approach to Address Fentanyl Overdose: Formative Study (e37483)</p> <p>Vanessa McMahan, Justine Arenander, Tim Matheson, Audrey Lambert, Sarah Brennan, Traci Green, Alexander Walley, Phillip Coffin. . . . .</p>	565
<p>Objective Monitoring of Facioscapulohumeral Dystrophy During Clinical Trials Using a Smartphone App and Wearables: Observational Study (e31775)</p> <p>Ghobad Maleki, Ahnjili Zhuparris, Ingrid Koopmans, Robert Doll, Nicoline Voet, Adam Cohen, Emilie van Brummelen, Geert Groeneveld, Joris De Maeyer. . . . .</p>	577
<p>A Versatile and Scalable Platform That Streamlines Data Collection for Patient-Centered Studies: Usability and Feasibility Study (e38579)</p> <p>Haley Huang, Sofia Aschettino, Nasim Lari, Ting-Hsuan Lee, Sarah Rosenberg, Xinyi Ng, Stella Muthuri, Anirudh Bakshi, Korrin Bishop, Hussein Ezzeldin. . . . .</p>	590
<p>Implementation of a Work-Related Asthma Screening Questionnaire in Clinical Settings: Multimethods Study (e37503)</p> <p>Madison MacKinnon, Max Moloney, Emma Bullock, Alison Morra, Teresa To, Catherine Lemiere, M Lougheed. . . . .</p>	607
<p>Digital Platform to Continuously Monitor Patients Using a Smartwatch: Preliminary Report (e40468)</p> <p>Kaio Bin, Lucas De Pretto, Fabio Sanchez, Linamara Battistella. . . . .</p>	621
<p>Using Wake-Up Tasks for Morning Behavior Change: Development and Usability Study (e39497)</p> <p>Kyue Oh, Jisu Ko, Jaemyung Shin, Minsam Ko. . . . .</p>	633
<p>Tough Talks Virtual Simulation HIV Disclosure Intervention for Young Men Who Have Sex With Men: Development and Usability Testing (e38354)</p> <p>Lisa Hightow-Weidman, Kathryn Muessig, Zach Soberano, Matthew Rosso, Andrew Currie, Margo Adams Larsen, Kelly Knudtson, Alyssa Vecchio. . . . .</p>	647
<p>Optimizing Health Coaching for Patients With Type 2 Diabetes Using Machine Learning: Model Development and Validation Study (e37838)</p> <p>Shuang Di, Jeremy Petch, Hertzger Gerstein, Ruoqing Zhu, Diana Sherifali. . . . .</p>	661
<p>Postoperative Outcomes of a Digital Rehabilitation Program After Total Knee Arthroplasty: Retrospective, Observational Feasibility Study (e40703)</p> <p>Mindy Hong, Joey Loeb, Manshu Yang, Jeannie Bailey. . . . .</p>	671

<b>mHealth-Supported Gender- and Culturally Sensitive Weight Loss Intervention for Hispanic Men With Overweight and Obesity: Single-Arm Pilot Study (e37637)</b>	
David Garcia, Luis Valdez, Benjamin Aceves, Melanie Bell, Brooke Rabe, Edgar Villavicencio, David Marrero, Forest Melton, Steven Hooker. . .	
6	7
<b>Feasibility of Measuring Screen Time, Activity, and Context Among Families With Preschoolers: Intensive Longitudinal Pilot Study (e40572)</b>	
Hannah Parker, Sarah Burkart, Layton Reesor-Oyer, Michal Smith, Roddrick Dugger, Lauren von Klinggraeff, R Weaver, Michael Beets, Bridget Armstrong. . . . .	692
<b>Extracurricular Humanism in Medicine Initiative and Medical Student Wellness: Retrospective Study (e37252)</b>	
Elizabeth Auckley, Jeff Barbee, Nicole Verbeck, Tracie McCambridge, Linda Stone, Jennifer Garvin. . . . .	708
<b>Exploring Motivations for COVID-19 Vaccination Among Black Young Adults in 3 Southern US States: Cross-sectional Study (e39144)</b>	
Marie Stoner, Erica Browne, David Tweedy, Audrey Pettifor, Allysha Maragh-Bass, Christina Toval, Elizabeth Tolley, Maria Comello, Kathryn Muessig, Henna Budhwani, Lisa Hightow-Weidman. . . . .	726
<b>The Effect of Persuasive Design on the Adoption of Exposure Notification Apps: Quantitative Study Based on COVID Alert (e34212)</b>	
Kiemute Oyibo, Plinio Morita. . . . .	737
<b>Circulating Illness and Changes in Thermometer Use Behavior: Series of Cross-sectional Analyses (e37509)</b>	
Jack Seifarth, Megan Pinaire, John Zicker, Inder Singh, Danielle Bloch. . . . .	765
<b>Optimizing the Acceptability, Adherence, and Inclusiveness of the COVID Radar Surveillance App: Qualitative Study Using Focus Groups, Thematic Content Analysis, and Usability Testing (e36003)</b>	
Bas Splinter, Nicholas Saadah, Niels Chavannes, Jessica Kiefte-de Jong, Jiska Aardoom. . . . .	779
<b>Satisfaction With Telemedicine in Patients With Orthopedic Trauma During the COVID-19 Lockdown: Interview Study (e35718)</b>	
Thomas Rauer, Julian Scherer, Pascal Stäubli, Jonas Gerber, Hans-Christoph Pape, Sandro-Michael Heining. . . . .	794
<b>The Impacts of Social Media Use and Online Racial Discrimination on Asian American Mental Health: Cross-sectional Survey in the United States During COVID-19 (e38589)</b>	
Alyan Layug, Samiksha Krishnamurthy, Rachel McKenzie, Bo Feng. . . . .	800
<b>Digital Storytelling Methods to Empower Young Black Adults in COVID-19 Vaccination Decision-Making: Feasibility Study and Demonstration (e38070)</b>	
Allysha Maragh-Bass, Maria Comello, Elizabeth Tolley, Darrell Stevens Jr, Jade Wilson, Christina Toval, Henna Budhwani, Lisa Hightow-Weidman. . . . .	817
<b>Discovering Long COVID Symptom Patterns: Association Rule Mining and Sentiment Analysis in Social Media Tweets (e37984)</b>	
Surani Matharaarachchi, Mike Domaratzki, Alan Katz, Saman Muthukumarana. . . . .	832
<b>Public Interest and Accessibility of Telehealth in Japan: Retrospective Analysis Using Google Trends and National Surveillance (e36525)</b>	
Takuya Kinoshita, Takehiro Matsumoto, Naota Taura, Tetsuya Usui, Nemu Matsuya, Mayumi Nishiguchi, Hozumi Horita, Kazuhiko Nakao. . .	
8	4
<b>Actions Speak Louder Than Words: Sentiment and Topic Analysis of COVID-19 Vaccination on Twitter and Vaccine Uptake (e37775)</b>	
Murooj Yousef, Timo Dietrich, Sharyn Rundle-Thiele. . . . .	856

---

Health Information Sourcing and Health Knowledge Quality: Repeated Cross-sectional Survey ( <a href="#">e39274</a> ) Elena Korshakova, Jessecae Marsh, Samantha Kleinberg. ....	879
--	-----

## Viewpoint

Collaborative Challenges of Multi-Cohort Projects in Pharmacogenetics—Why Time Is Essential for Meaningful Collaborations ( <a href="#">e36759</a> ) Filippo Franchini, Katharina Kusejko, Catia Marzolini, Christoph Tellenbach, Simona Rossi, Susanne Stampf, Michael Koller, Jivko Stoyanov, Burkhard Möller, Alexander Leichtle. ....	717
--	-----

## Corrigenda and Addenda

Correction: Using the Transformative Storytelling Technique to Generate Empowering Narratives for Informal Caregivers: Semistructured Interviews, Thematic Analysis, and Method Demonstration ( <a href="#">e42323</a> ) Milica Petrovic, Silvia Bonanno, Marta Landoni, Chiara Ionio, Mariët Hagedoorn, Andrea Gaggioli. ....	830
---	-----

Original Paper

# Measuring Daily Activity Rhythms in Young Adults at Risk of Affective Instability Using Passively Collected Smartphone Data: Observational Study

Benny Ren<sup>1</sup>, MSc; Cedric Huchuan Xia<sup>2,3</sup>, MD, PhD; Philip Gehrman<sup>3,4</sup>, PhD; Ian Barnett<sup>1</sup>, PhD; Theodore Satterthwaite<sup>2,3</sup>, MD

<sup>1</sup>Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

<sup>2</sup>Penn Lifespan Informatics and Neuroimaging Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

<sup>3</sup>Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States

<sup>4</sup>Michael J Crescenzo VA Medical Center, Philadelphia, PA, United States

**Corresponding Author:**

Benny Ren, MSc

Department of Biostatistics, Epidemiology, and Informatics

University of Pennsylvania Perelman School of Medicine

1st floor Blockley Hall

423 Guardian Drive

Philadelphia, PA, 19104

United States

Phone: 1 9177676698

Email: [bennyren@penmedicine.upenn.edu](mailto:bennyren@penmedicine.upenn.edu)

## Abstract

**Background:** Irregularities in circadian rhythms have been associated with adverse health outcomes. The regularity of rhythms can be quantified using passively collected smartphone data to provide clinically relevant biomarkers of routine.

**Objective:** This study aims to develop a metric to quantify the regularity of activity rhythms and explore the relationship between routine and mood, as well as demographic covariates, in an outpatient psychiatric cohort.

**Methods:** Passively sensed smartphone data from a cohort of 38 young adults from the Penn or Children's Hospital of Philadelphia Lifespan Brain Institute and Outpatient Psychiatry Clinic at the University of Pennsylvania were fitted with 2-state continuous-time hidden Markov models representing active and resting states. The regularity of routine was modeled as the hour-of-the-day random effects on the probability of state transition (ie, the association between the hour-of-the-day and state membership). A regularity score, Activity Rhythm Metric, was calculated from the continuous-time hidden Markov models and regressed on clinical and demographic covariates.

**Results:** Regular activity rhythms were associated with longer sleep durations ( $P=.009$ ), older age ( $P=.001$ ), and mood ( $P=.049$ ).

**Conclusions:** Passively sensed Activity Rhythm Metrics are an alternative to existing metrics but do not require burdensome survey-based assessments. Low-burden, passively sensed metrics based on smartphone data are promising and scalable alternatives to traditional measurements.

(*JMIR Form Res* 2022;6(9):e33890) doi:[10.2196/33890](https://doi.org/10.2196/33890)

**KEYWORDS**

mobile health; mHealth; hidden Markov model; mental health; circadian rhythm; mobile phone

## Introduction

**Background**

The proliferation of smartphone use in mobile health (mHealth) research has resulted in a wealth of longitudinal data capable of quantifying human behaviors pertinent to the study of mental

health [1-3]. Smartphones collect a wide variety of sensor data, ranging from accelerometer and geolocation (ie, GPS) data to screen time and social interactions, which are increasingly being used as digital biomarkers of behavior in a variety of contexts [4-7]. The continuous collection of this wealth of data enables us to study an individual's pattern of behavior across the course

of each day. Many behaviors show a diurnal rhythm, an observed 24-hour periodic pattern, some of which are measurable through digital biomarker data [6,8]. These rhythms reflect endogenous physiological circadian processes related to many clinically relevant outcomes [9]. A wide range of physiological processes follows a circadian rhythm [10-12], such as cardiometabolic function and gene expression [13-15]. Certain immunological processes and drug efficacy are sensitive to specific points in the circadian cycle, highlighting the need to understand the role of circadian rhythms from a pharmacodynamics perspective [16,17]. In addition, disruptions in rest-activity cycles have been associated with adverse outcomes in posttraumatic stress and affective disorder studies [18,19]. This underscores the need to meaningfully quantify circadian rhythms in ecological contexts, such as the assessment of diurnal rhythms, as reflected in smartphone use data.

Before mobile devices were used to gather high-frequency ecological momentary assessment (EMA) and continuous streams of sensor data collection, information obtained from diaries and surveys was used for scoring the regularity of diurnal activities [20,21]. For example, Social Rhythm Metric (SRM) uses daily administered diaries to record the timing of routine activities (eg, getting out of bed, eating lunch, and starting work). To score highly on the SRM, one must consistently perform these activities close to the same time of day for most records, such as consistently waking up at the same time every day. Subsequently, high scores can be interpreted as greater regularity in routine or rhythm and provide a useful quantification of the regularity of rhythms, which can be used to study clinical outcomes. Disruptions to the regularity of rhythms have been associated with psychiatric disorders (eg, bipolar disorders [22-24], anxiety disorders [25], depression [12,26], mood or affective disorders [27,28], posttraumatic stress disorder [27,29], and substance-related disorders [30,31]). Measures of regularity are often markers of treatment efficacy in social rhythm therapies aimed at improving mental health [27]. Despite the utility of diary-based scores in the studies of psychopathology, manual data collection could be burdensome, susceptible to self-report bias, recall bias, nonresponse bias, and experience degradation over time because of the loss of engagement with the participant [1,32-34].

The potential use of smartphones for long-term data collection is made possible by ensuring a low burden on the participant, particularly through sensor data that can be collected passively (ie, without requiring active participation or input from the user) [35,36]. As irregular circadian rhythms are associated with a range of psychiatric disorders, particularly affective disorders [28,37,38], the logical next step is to use smartphone sensor data to quantify regularity in diurnal rhythms and then identify correlated clinically relevant outcomes [4]. If passively collected smartphone sensor data can be used to inform clinically relevant behaviors and symptoms, this would provide an effective and low-burden approach to mental health assessment and treatment monitoring [39].

## Objectives

Many models have been proposed to study circadian or diurnal rhythms based on passively collected sensor data, ranging from

simple rule-based analyses to deep learning models [18,40-42]. Longitudinal data collected in mHealth studies can be viewed as a multivariate time series and subsequently often draw upon a variety of longitudinal and stochastic models [43-46]. A focal point of mHealth studies is to model rest-activity cycles or other categorical outcomes [47-50]. Modeling the dichotomy of rest-active states often simplifies clustering [8,51,52] or classification [5,52,53] problems where rhythmic effects can be modeled with harmonics [45,54-56]. For our purposes, to model transitions between different circadian biological states over the course of the day, we modeled these data using hidden Markov models (HMMs). In this framework, we captured the circadian nature of behaviors, as measured using smartphone sensor data, through the use of hour-of-the-day random effects [57-59].

Despite the potential of mHealth data, statistical models that translate these data into interpretable measures of diurnal rhythms and markers to manage mental health are an active area of research [27,39]. In the time series literature, random effects have been used to model seasonality or other periodic effects [59,60]. Our continuous-time HMM (CT-HMM) transitions between rest-activity cycles use individual-specific random intercepts for hours of the day (eg, 12 AM to 1 AM and 1 AM to 2 AM) to allow for personalized patterns of diurnal activity [61]. By fitting this model to each individual separately, we were able to quantify the regularity of activity rhythms or routines and determine how this strength of routine is correlated with a variety of demographic and clinical outcomes. Finally, we developed a novel score to gauge the regularity of activity rhythms and determine how this score correlates with self-reported sleep-related outcomes and other characteristics in a sample of adolescents with or at risk of affective instability.

## Methods

### Participants

A sample of 41 adolescents and young adults (28/41, 68% female participants) aged 17 to 30 (mean 23.4, SD 3.5) years were enrolled as part of a study on affective instability in youth. Participants were recruited via the Penn or Children's Hospital of Philadelphia Lifespan Brain Institute or through the Outpatient Psychiatry Clinic at the University of Pennsylvania [62]. Of these 41 participants, 38 (92%) met the criteria for an Axis I psychiatric diagnosis based on a semistructured clinical interview, and 33 (80%) met the criteria for >1 disorder [63]. In addition, 39% (16/41) of participants met the criteria for a personality disorder based on an assessment with the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders-4 Axis II Personality Disorders (Tables 1 and 2) [63]. As a secondary analysis, we used data collected from a prior study. Subsequently, the availability of new clinical data was a limitation of our retrospective study design. Although we had digital EMA and passive sensor data for all participants, baseline measurements such as the Pittsburgh Sleep Quality Index (PSQI) and psychiatric diagnoses were not available for all participants.



**Table 1.** Psychiatric diagnoses of participants (N=41).

Diagnosis	Participants, n (%)
<b>Axis I diagnosis</b>	
No diagnosis	3 (7)
<b>Diagnosis</b>	38 (93)
Major depressive disorder	24 (59)
Bipolar disorder	4 (10)
Depressive disorder NOS <sup>a</sup>	1 (2)
Mood disorder NOS	1 (2)
Generalized anxiety disorder	14 (34)
Posttraumatic stress disorder	14 (34)
Social phobia	12 (29)
Obsessive-compulsive disorder	11 (27)
Panic disorder	5 (12)
Anxiety disorder NOS	2 (5)
Attention-deficit or hyperactivity disorder	6 (15)
Schizoaffective disorder	1 (2)
Substance-related disorders	18 (44)
<b>Axis II diagnosis</b>	
No diagnosis	7 (17)
<b>Diagnosis</b>	16 (39)
Borderline personality disorder	12 (29)
Personality disorder NOS	4 (10)

<sup>a</sup>NOS: not otherwise specified.

**Table 2.** Baseline demographic and clinical characteristics (N=41).

Characteristics	Values
Sex (female), n (%)	28 (68)
Age (years), mean (SD)	23.4 (3.5)
Hours of Beiwe sensor data (screen-on and accelerometer), mean (SD)	1724 (753)
Beck Depression Inventory scores, mean (SD)	6.96 (8.45)
<b>Beiwe ecological momentary assessment, mean (SD)</b>	
“About how many hours did you actually sleep?”	7.43 (1.03)
“About what time did you go to bed last night, regardless of the time you actually fell asleep?”	11:28 PM (1.92 hours)
“What time did you wake up?”	7:41 AM (2.17 hours)
“How happy versus sad do you feel right now? (1-Very cheerful or happy, 2, 3, 4, 5, 6, 7-Very sad or depressed or unhappy)”	3.12 (1.5)

## Ethical Considerations

All participants provided informed consent for all study procedures. For minors, the parents or guardians, in addition to the minors, provided informed consent. This study was approved by the Institutional Review Board of the University of Pennsylvania (828424).

## Data Acquisition

From the 41 participants, 2972 person-days of sensor data, including accelerometer measures (meters per second squared) for the axes, were obtained from participant smartphones through the Beiwe app, a research platform developed by the Onnela Lab at the Harvard TH Chan School of Public Health [64]. Screen-on events for Android devices were recorded, whereas screen-unlock events for iOS devices were acquired through Beiwe; however, we have referred to both as *screen-on*

*events* in this paper for simplicity. Every morning, participants were also asked about their mood and sleep patterns and quality from the night before via self-report prompts delivered by Beiwe. These questions included sleep duration in hours (“About how many hours did you actually sleep?”), time to sleep (“About what time did you go to bed last night, regardless of the time you actually fell asleep?”), and time to wake (“What time did you wake up?”) and were obtained using self-report questionnaires. The possible time-to-sleep and time-to-wake responses were limited to hour-long intervals. Participants were asked to rank their mood with the following question: “How happy versus sad do you feel right now? (1-Very cheerful or happy, 2, 3, 4, 5, 6, 7-Very sad or depressed or unhappy).” A summary of the demographic and EMA covariates is provided in [Table 2](#). In addition to these questions administered through smartphones, additional measurements were collected at baseline, including the PSQI and the Beck Depression Inventory (BDI) scores [65,66].

### Data Processing

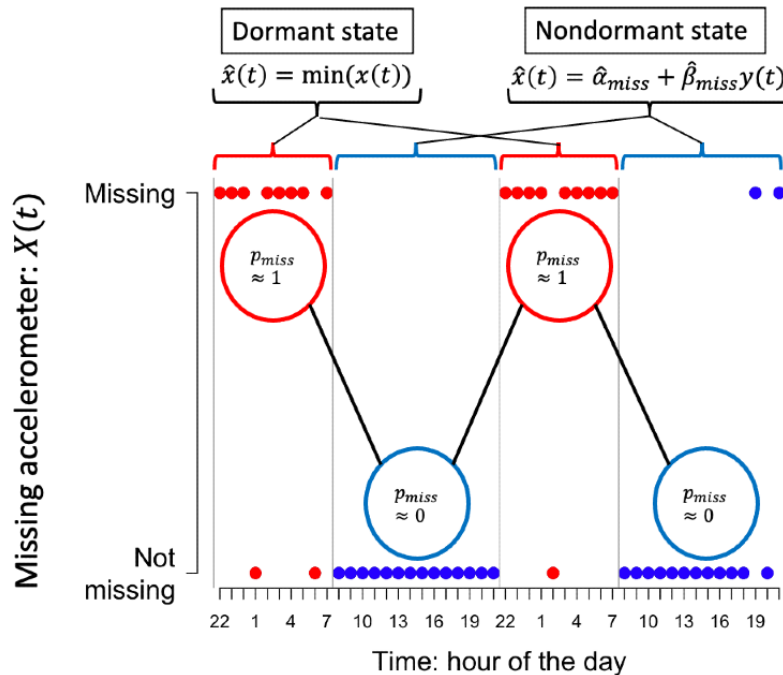
Our analysis goals are 2-fold: (1) to use smartphone sensors and activity data to quantify the strength of each participant’s activity rhythm or routine and subsequently (2) to test for significant associations between demographic variables or self-reported mood outcomes and the strength of activity rhythm or routine. Given our first goal of modeling a participant’s activity rhythm, we leveraged active smartphone use to provide an indicator of activity over the course of the day. From accelerometer data, hourly features were calculated to reflect the magnitude of movement of the smartphone. For each hour of the day, labeled by the time at the end of hour  $t$ ,  $X(t)$  is the mean of the magnitude of phone acceleration over the course of that hour. In addition,  $Y(t)$  is the screen-on count over the course of the hour. However, the periods of dormancy, where  $X(t)$  was unobserved, because of user- and device-related factors such as the phone being powered off, having no cell signal, or

being in airplane mode required accelerometer features to be imputed.

By considering the characteristics of screen-on events and accelerometer data, we designed a data imputation procedure guided by domain knowledge. Periods of dormancy usually align with periods of low phone use such as night-time hours, have a greater probability of missing accelerometer data  $X(t)$ , and were identified using a 2-state hidden semi-Markov model with Bernoulli state-dependent distributions [67]. For example, if accelerometer data are seldom missing over a given window of time and there are many screen-on events over the same period, it is likely that there was significant accelerometer activity despite being missing. Here, screen-on events can be used to impute accelerometer data. In contrast, if accelerometer data are missing and there are no screen-on events, then it is likely that the phone was in a state of dormancy with low accelerometer magnitudes. The periods of dormancy correspond to the 2 latent states in our hidden semi-Markov model imputation. Missing mean accelerometer magnitudes from dormant periods were imputed using the minimum (excluding outliers) mean accelerometer magnitudes  $X(t)$ , whereas missing data assigned to the nondormant state were imputed by regressing  $X(t)$  on  $Y(t)$  over all hours where data were completely observed. Ultimately, this led to an imputed  $X(t)$ , which we used in the following analyses, with a diagram of the imputation procedure outlined in [Figure 1](#).

The final source of noise in the data was user error, which often occurred when answering questionnaires, such as accidentally selecting 8 PM for wake-up time instead of 8 AM. To avoid these user errors, bedtimes were automatically corrected to be between 5 PM and 4 AM, and wake-up times were corrected to be between 4 AM and 3 PM when such extreme discrepancies occurred that resulted in unrealistic sleep durations that were almost certainly the result of measurement error.

**Figure 1.** Hidden semi-Markov model to identify periods of dormancy and missing data imputation. First, a hidden semi-Markov was used to identify periods of dormancy, which were characterized by a high proportion of missing accelerometer data, often associated with night-time hours. Hidden semi-Markov allows for a geometric dwell time distribution with Bernoulli outcomes to model the proportion of missingness over consecutive hours. Second, missing mean accelerometer magnitudes  $X(t)$  for dormant states were imputed using the minimum (excluding outliers) mean accelerometer magnitudes  $X(t)$ , whereas missing data for nondormant states were imputed using linear regression.



**Activity Rhythm Modeling**

Stochastic models are often used to study longitudinal data sets such as the data generated by smartphones, and we opted to use a continuous-time Markov chain framework with the addition of random intercepts representing each hour of the day to model activity rhythms for each participant separately. This choice was driven in part by its ability to account for missing data for which many harmonic analyses were not designed [68]. If an hour of the day has a large random intercept, it represents a higher probability of active phone use during that hour relative to other hours of the day for a specific participant. With this interpretation in mind, a participant with a strong activity rhythm will have hour-of-the-day random intercepts with a large magnitude or, equivalently, with high variability. In addition, the variances of the random intercepts are the test statistics of a mixed-effect ANOVA, where the null hypothesis is that the hour of the day has no effect on the state transition. Phone use, binned into hour-long intervals of activity or rest, was fitted with a 2-state continuous-time Markov chain. By modeling transition rates with an exponential proportional hazard (PH) regression with time-varying covariates, we treated the state labels as latent variables, which correspond to an HMM.

In our HMM, screen-on counts,  $Y(t)$ , were characterized by a mixture of 2 state-dependent Poisson distributions with a rest state,  $C(t)=2$ , where  $E[Y(t)|C(t)=2] \approx 0$ , and an active state,  $C(t)=1$ , where  $E[Y(t)|C(t)=1] > 0$ . We also incorporated accelerator magnitude averaged over the hour,  $X(t)$ , and hour-of-the-day random intercepts (or frailties) in an exponential PH regression used to estimate rates of transitioning from the rest-to-active and active-to-rest states. Random intercepts can be viewed as a penalized effect that the hour of the day has on

transition rates and can be interpreted as the activity rhythm. Using our rates and event times, we use Kolmogorov equations to estimate transition probability matrices and construct a 2-state mixed CT-HMM [69]. The transition rates are as follows:

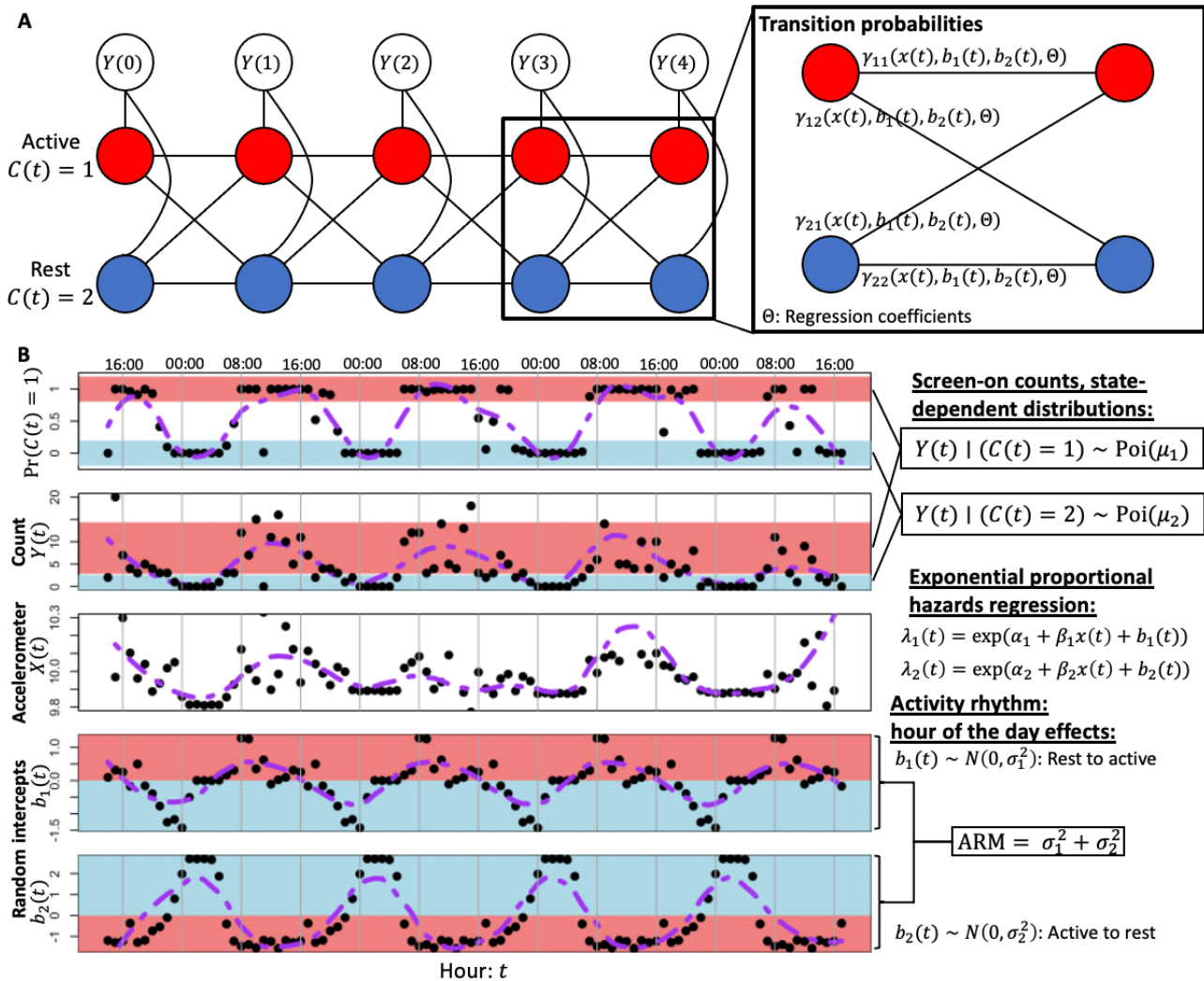
$$\text{Rest to active: } \lambda_1(t) = \exp(\alpha_1 + \beta_1 \times x[t] + b_1[t]), \text{ with } b_1(t) \sim N(0, \sigma_1^2) \quad (1)$$

$$\text{Active to rest: } \lambda_2(t) = \exp(\alpha_2 + \beta_2 \times x[t] + b_2[t]), \text{ with } b_2(t) \sim N(0, \sigma_2^2) \quad (2)$$

where  $b_1(t)$  and  $b_2(t)$  are random intercepts for the hour of the day. Subsequently, the corresponding transition rate matrices,  $Q(t)$ , are functions of  $\lambda_1(t)$  and  $\lambda_2(t)$ . The transition probability matrices are given by the matrix exponential  $\Gamma(t) = e^{Q(t)}$  as the event times are 1-hour increments (Figure 2). In cases where periods of consecutive missing accelerometer data continue over 24 consecutive hours, this constitutes a sufficiently long period of missing data that requires splitting the HMM into 2 segments on either side of the missing interval, where the likelihood of the multiple HMMs can be treated as independent and multiplied together during parameter estimation. The CT-HMM is fitted with the expectation-maximization (EM) algorithm by iteratively solving  $\Theta = (\alpha_1, \beta_1, \alpha_2, \beta_2), b_1(t), b_2(t), \sigma_1^2, \sigma_2^2$  and  $Pr \times (C[t]=i, C[t+1]=j)$  [70-72]. A high frequency of missing data can result in an identifiability problem when fitting the HMM; for example, when an individual seldom uses their phone and the 2 states of the HMM become indistinguishable as the data may not reflect the daily differences in activity rhythms. Seldom phone use results in low screen-on counts and missing accelerometer data, leading to unimodal data streams, even after imputation, where daily differences are difficult to identify. Following this procedure, 7% (3/41) of

individuals were omitted from the analysis as their EM algorithms failed to converge because of the much higher than normal frequency of missing data, resulting in a final sample size of 38.

**Figure 2.** Hidden Markov model framework and the ARM. (A) Graph of the 2-state continuous-time hidden Markov model used to model phone engagement via screen-on counts. Active states are characterized by expected screen-on counts >0, and the rest states are characterized by expected screen-on counts close to 0. Transition rates between rest and active states are modeled with exponential proportional hazard regressions by using 24 hours of the day as random intercepts. (B) Hourly state membership probabilities, state-dependent distribution counts, accelerometer norms, and random intercepts for the hour of the day. We controlled for accelerometer activity in the regression models while estimating activity rhythm with random intercepts. Accelerometer activity is positively correlated with rest-to-active transitions and negatively correlated with active-to-rest transitions. Large hour-of-the-day effects correspond to a regular activity rhythm and resulted in a large ARM. ARM: Activity Rhythm Metric.



### Comparing Activity Rhythms, Self-reported Sleep, and Depression-Related Variables

After modeling each participant’s activity rhythm via the CT-HMM—random intercepts corresponding to the effect of the hour of the day on the likelihood of rest versus activity—we naturally expect regularity in activity profiles to correspond with large values of the variance of the random intercepts as quantified by  $\sigma_1^2$  and  $\sigma_2^2$ . In other words, if a person tends to be active or at rest during the same hours of the day routinely, then the restful hours will have very low random intercepts, with active hours having high random intercepts. Under the mixed-effects ANOVA, the large random intercept variances correspond to rejecting the null hypothesis that the hour of the day has no effect on state transition. This large discrepancy in random intercepts between different hours of the day manifests

as large values of  $\sigma_1^2$  and  $\sigma_2^2$ . Thus, for each individual, we fit a CT-HMM and sum over these 2 variance terms to obtain what we define as Activity Rhythm Metric (ARM):  $\text{ARM} = \sigma_1^2 + \sigma_2^2$  (Figure 2).

With quantification of the strength of each participant’s activity rhythm or routine through the ARM, we validate the ARM as a measure of daily rhythm by treating it as the outcome in a linear regression to test for associations with the following self-reported sleep-related covariates: mean nightly sleep duration (mean response of “About how many hours did you actually sleep?”), the baseline sleep duration component of the PSQI, time-to-sleep SD (SD of “About what time did you go to bed last night, regardless of the time you actually fell asleep?”), and time-to-wake SD (SD of “What time did you wake up?”). We would expect a higher ARM to correspond

with a stronger routine and therefore with small time-to-sleep or time-to-wake SDs and longer mean sleep duration. For each sleep-related covariate alone, we fit linear regression models controlling for age and sex and compared it with a null model with only age and sex, using a likelihood ratio test (LRT) to test for the association in a 2-sided alternative hypothesis. We used the mean mood response from the Beiwe questionnaire (mean response of “How happy versus sad do you feel right now?”) and BDI as a depression-related measure, with higher mean values corresponding to severe depression. For each depression-related covariate, we tested for an association with the ARM by using the same LRT framework. In addition, we tested the association between age or sex and the ARM while controlling for the others by using the LRT.

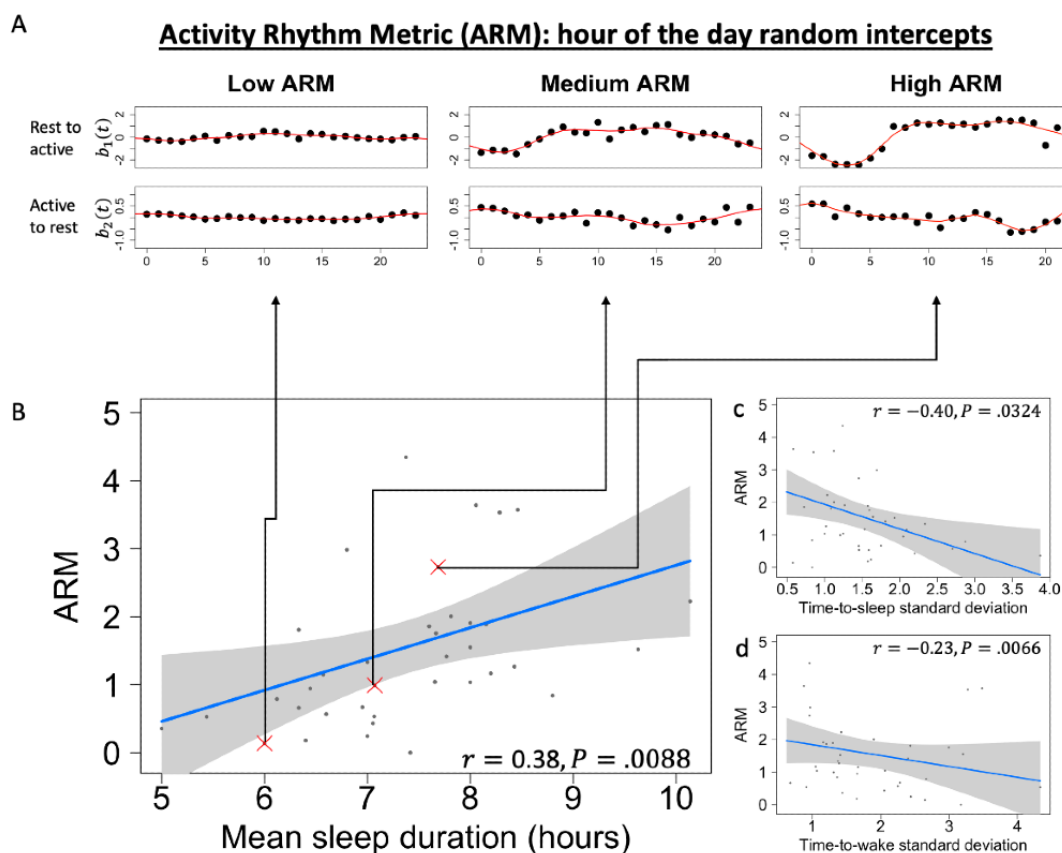
## Results

### Relationship With Sleep Duration

We found that all sleep-related measures, namely, mean nightly sleep duration, the baseline sleep duration component of the

PSQI, time-to-sleep SD, and time-to-wake SD, were marginally significantly ( $P < .05$ ) associated with the ARM (Figure 3). Association tests included only individuals with corresponding self-reported outcomes. Notably, continuous daily administration of sleep surveys may increase the patient burden in a study, highlighting the advantages of passive data collection methods. We found that individuals with a higher ARM are more likely to have longer sleep duration, with an hourly increase in mean sleep duration corresponding to a 0.4 increase in the ARM. The mean nightly sleep duration from the Beiwe questionnaires (“About how many hours did you actually sleep?”) captures the same information as the baseline sleep duration component of the PSQI, which was also significantly ( $P < .001$ ) associated with the ARM and had the same direction of effect. Considering sleep duration as a component of sleep quality, these findings suggest that the ARM was positively correlated with sleep quality. In other words, a stronger and more consistent routine, as measured passively through smartphone use, corresponds to better sleep quality.

**Figure 3.** The Activity Rhythm Metric and self-reported sleep. (A) Visualization of 3 example participants with hour-of-the-day random intercepts mapping to low, medium, and high Activity Rhythm Metrics. Random intercepts are the hour-of-the-day effect on state transition: low or high Activity Rhythm Metric score participants have low or high variability in random intercepts. (B) Linear regression models were fit between the Activity Rhythm Metric and sleep-related outcomes across all participants using a likelihood ratio test for an association with each sleep measure while controlling for age and sex. Higher Activity Rhythm Metrics are associated with a longer mean duration of sleep ( $r=0.38$ ). (C and D) Higher Activity Rhythm Metrics are associated with less variability in self-reported time-to-sleep ( $r=-0.40$ ) and time-to-wake ( $r=-0.23$ ) responses.



### Relationship With Self-reported Sleep and Wake Times

The questionnaires captured by Beiwe through smartphones prompted participants to report time to sleep and time to wake,

which have been used to measure the regularity of activity rhythms in other contexts [73]. We expected that the high variance in self-reported time-to-sleep and time-to-wake responses would correspond to irregular routines. This finding

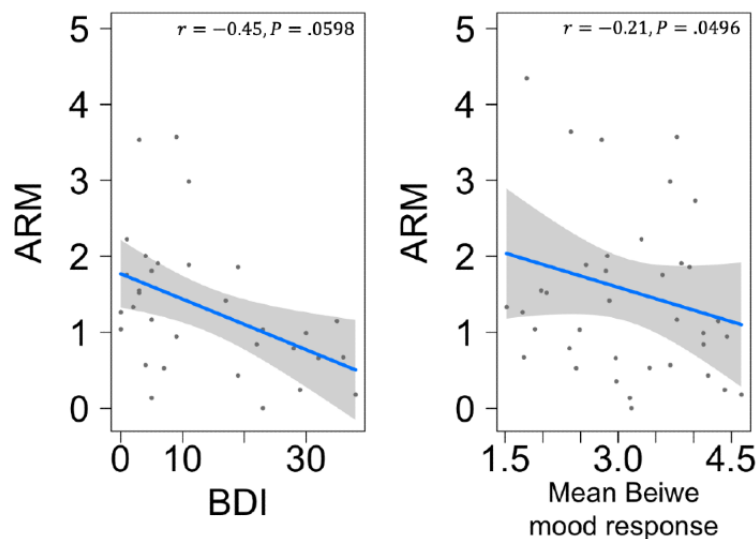
is consistent with the relationship between the ARM and variability in self-reported sleep or wake timing. In particular, we found that the ARM is significantly correlated with the SDs of time-to-sleep (“About what time did you go to bed last night, regardless of the time you actually fell asleep?”) and time-to-wake (“What time did you wake up?”) responses (Figure 3). A unit increase in time-to-sleep SD corresponds to a 0.52 decrease in the ARM, and a unit increase in time-to-wake SD corresponds to a 0.48 decrease in the ARM. Thus, our proposed ARM measure passively captures many of the same routine-related signals as traditional survey-based metrics while avoiding the high burden of daily self-reporting, which is otherwise necessary to collect data on the variability of the sleep schedule.

### Relationship With BDI and Self-reported Mood

In a variety of clinical populations, there is evidence of a relationship between depression-related metrics and irregular

routines, where the quantification of regular rhythms can be used to assess treatment efficacy [27]. In line with this, we found that the ARM was negatively correlated with the BDI (Figure 4); that is, irregular activity rhythms were associated with higher BDI. Of note, only the response to the Beiwe mood question (“How happy versus sad do you feel right now?”) was marginally associated with the ARM ( $P=.049$ ); however, the direction of the associations was intuitive, albeit borderline significant (Figure 4). Of note, 15% (6/41) of additional participants had missing BDI scores. The average ARM for the participants with missing BDI was 2.86, whereas it was much lower (1.31) for participants who had recorded the BDI; it is likely that the missing BDI information, if observed, would help increase the precision of the association and increase its strength. The difference between the average BDI values of these 2 groups suggests that informative missingness may have diminished statistical power.

**Figure 4.** Activity rhythm and its relationship with mood and depression. In total, 2 mood or depression-related self-reported outcomes were compared with the ARM: (1) The BDI was recorded for 32 participants, and (2) the mean response from the Beiwe smartphone mood question: “How happy versus sad do you feel right now? (1-Very cheerful or happy, 2, 3, 4, 5, 6, 7-Very sad or depressed or unhappy).” Linear regression models were fit by using a likelihood ratio test while controlling for age and sex. ARM and BDI have a negative correlation ( $r=-0.45, P=.06$ ). ARM and mean response from the Beiwe mood survey have a negative correlation ( $r=-0.21, P=.049$ ). ARM: Activity Rhythm Metric; BDI: Beck Depression Inventory.

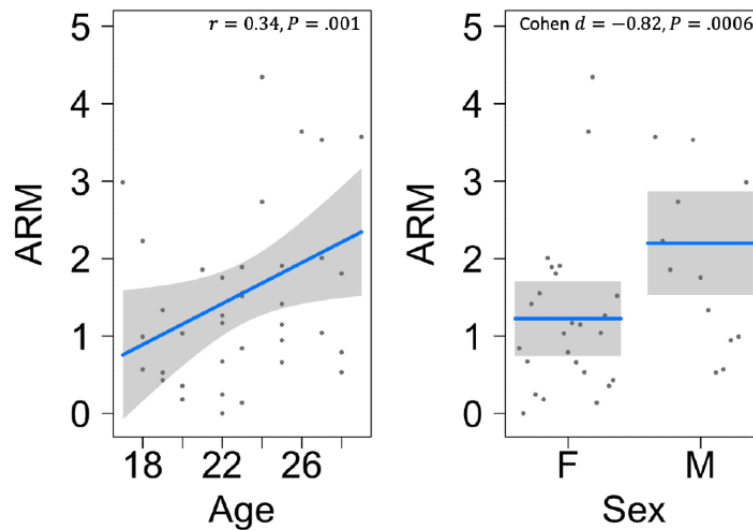


### Relationship With Age and Sex

In addition to the ARM being associated with sleep-related measures of duration and variability in time-to-sleep and time-to-wake responses, we found that the ARM was associated with age in a manner similar to previous regularity of rhythm studies, although previous studies examined different study populations [73,74]. We found a positive correlation between

age and the ARM ( $P=.001$ ); a year increase in age corresponds with a 0.18 increase in the ARM, meaning that older individuals tend to have more regular activity rhythms and routines (Figure 5). In addition, our analysis of the ARM suggests that there is a significant ( $P<.001$ ) sex-based difference in the regularity of a participant’s activity rhythm (Figure 5). We expect a 1.31 increase in the ARM of male participants in our sample relative to that of female participants.

**Figure 5.** Activity rhythm and its relationship with age or sex. Linear regression models were fit by using a likelihood ratio test for age or sex each while controlling for the other covariate. Age is positively correlated with ARM ( $r=0.34$ ,  $P=.001$ ), and male participants have higher ARMs (Cohen  $d=-0.82$ ,  $P<.001$ ). ARM: Activity Rhythm Metric.



## Discussion

### Principal Findings

The regularity of daily routine, as measured through the ARM—a quantification of routine based solely on passively collected smartphone data—was found to be significantly associated with a variety of demographic, mood, and sleep-related measures. We developed a CT-HMM that allows for an hour-of-the-day effect on state membership (active vs rest). Using the variance of the hour-of-the-day random intercepts to represent the strength of the routine, we constructed an ARM and found it to be associated with the SD in self-reported time-to-bed and time-to-wake responses on a night-to-night basis. These findings validate the ARM as a quantification of the strength of routine, which can be used as an outcome in studies aimed at improving mental health by increasing regularity in routine. Furthermore, the ability to calculate the ARM using only passively collected smartphone data provides a crucial advantage relative to the traditional reliance on self-report to dynamically quantify routine. This can provide a low-burden alternative that can easily be deployed at scale, even in studies with long follow-up durations, as passively collected data are not susceptible to the same survey fatigue, which makes long-term follow-up a challenge in studies that rely heavily on self-report.

### Comparison With Prior Work

The direction of associations for the ARM is aligned with existing metrics such as the SRM. The SRM is a diary-based metric that has helped inspire our ARM definition and approach, where both are calculated from intraindividual routine variations, and higher scores correspond to regular routines [20,21]. The ARM evaluates active or rest states timing akin to calculating sleep or wake variability in time from data obtained through a survey or diary. Both the ARM and SRM are highly influenced by variances in the timing of habitual daily behaviors, such as time to sleep and time to wake, where a high variance corresponds to a low score. Our findings showed that a low

ARM is associated with high variability in time-to-sleep and time-to-wake self-reported responses, indicating that the ARMs are correlated with information that would otherwise be obtained through diary-based metrics.

Our results, which link the ARM to sleep duration, reinforce some findings derived from diary-based methods. Monk et al [73,75] found that the SRM was negatively correlated with better sleep quality, as measured by the PSQI. Similar to the SRM, we found that the ARM tends to increase with age, with older participants tending to have greater regularity in their routines [73,74]. Our analysis also showed greater regularity in activity rhythms in male participants, which aligns with the findings by Monk et al [67], despite some studies showing conflicting data on sex-based differences in the SRM [20,73,76,77]. Although the ARM is inspired by the SRM, we noted that the ARM fundamentally represented a narrower scope than the SRM by only focusing on a person's activity and with no direct measurement of social behavior.

We expected that the ARM would tend to decrease with higher depressive symptoms, reflecting the established relationship between stronger routines and milder depression [27]. Although the same direction of effect was shown in our sample (Figure 4), we lacked the sample size to achieve more than marginal statistical significance in this association. Similar psychiatric studies had access to larger cohorts of participants and healthy controls, which we lacked in this study, and would greatly improve the statistical power [20,26]. An important next step is to repeat these analyses with a larger sample size to validate which relationships hold. Consequently, the potential to quantify routines using only passively collected data may be an informative and actionable digital biomarker with respect to clinically relevant outcomes such as depression.

### Limitations

Our analysis has limitations typical of a retrospective study—a secondary analysis with limited data. The absence of a control group could have affected the statistical power. For example,

in a case-control study, we expected the control group to have high ARMs and low BDI scores. Data from a control group would be high leverage points in regression analysis in certain situations and would increase the effect size. In addition, this study was limited to a small age range. An in-depth analysis of regularity was not anticipated during the initial recruitment of the study cohort; subsequently, diary-based regularity metrics were not available for a direct comparison with ARMs. In summary, our analysis explored connections between routine activity rhythms and several clinical covariates; however, the validation of relationships and generalization to broader populations are left to future studies with prospective designs. In addition, differences between device hardware and operating systems could introduce heterogeneity in the data and should be accounted for in mHealth studies. For example, iOS acceleration is normalized by the g-force constant, whereas Android acceleration is not. Our individual-specific PH regressions, as discrete class models, were not affected by a scale difference in the covariates, and the g-force constant was absorbed by the coefficient estimate.

Although the use of passively collected smartphone data successfully reduces participant burden, there are some hurdles in this type of data collection. In certain cases, many clinical populations [78], including ours, are willing to share sensor data despite privacy concerns. In addition, sharing ARM-like summary statistics rather than the entire collection of sensor data may attenuate the privacy concerns of clinicians, allowing studies to increase the sample size. However, accounting for the missing data of various missing mechanisms remains paramount, and there is a lack of methods available to handle missingness when it is associated with the outcome of interest (missing not-at-random). Although we proposed a simple domain-based approach that takes into account missingness because of a lack of phone engagement induced by diurnal patterns, additional sensitivity analyses using different imputation procedures are necessary to completely understand the effect of missing data in mHealth studies.

In addition, identifiability is an important concern in mixed-effect modeling. A sufficient sample size for each hour must be available for modeling random intercepts. HMMs, which are models with many parameters, require large sample sizes or strong signals for parameter estimation. Although our EM algorithm failing to converge indicates an identifiability concern with the underlying data, other criteria may be used for the explorative analysis of data before HMM fitting. In participants with limited data, the hour-of-the-day effects must be pronounced to fit the random intercept model. In other words, sufficient evidence is needed to detect hour-of-the-day effects, either adequate sample size or the strength of association between hour-of-the-day and state membership. In our case, a lack of engagement with the study phone leading to unimodal data streams is an important consideration related to model fit but is addressed by our failure of convergence criterion. Proper use of study phones is an important prerequisite for mHealth studies, and exploring steps for filtering out individuals based on data quality is an important component of the mHealth study design. Data quality, both handling of missing data and proper use of the study device, remains a paramount concern in

mHealth studies. As a result, we look to evaluate the many relationships uncovered by our analyses in future work involving different study cohorts and different methods for handling missing data.

### Potential Future Directions

Our novel CT-HMM framework could easily be adapted to incorporate additional information to improve its ability to quantify diurnal activity patterns and routines. For example, we can extend the univariate outcome,  $Y(t)$ , to a multivariate joint probability distribution that also considers longitudinal GPS data. Although we modeled a parsimonious representation of daily routines, a common prompt in SRM diaries is the time of starting work, which can be ascertained from GPS location. By incorporating GPS data and increasing the number of states in our CT-HMM beyond the 2 rest or active states used in this study, we can model a wider range of routine behaviors and participant states.

Alternative HMM formulations that allow latent states to represent more than just rest or active states, such as symptom severity, would make the expected timing of state transitions clinically relevant. Many mixed-effect HMMs use a logit link or logistic regression to model transitions between states, where coefficients can be interpreted as odds ratios [60,61]. We elected to use a PH model, where the coefficients can be interpreted as hazard ratios. Although the interpretation of the signs of the coefficients is similar across both models, under the PH model, the expected event time (or time until state transition) can be calculated as  $1/\lambda(t)$ . This expected event time is intuitively important as it allows for the prediction of state changes in an individual, which can be used to prompt a mHealth intervention in the context of an HMM framework where latent states could represent, for example, manic, or depressed states in an individual with bipolar disorder.

### Conclusions

The previous generation of diary-based metrics comparable with the ARM are limited by self-report, which requires a high burden on the participant, underscoring the potential of mHealth solutions. However, the identifiability of complex models for mHealth data should be taken into consideration during study design. We estimated each participant's activity rhythm and corresponding strength of routine by calculating participant-specific hour-of-the-day random intercepts in a novel CT-HMM modeling framework that dictated consistency in phone activity over the course of the day for each participant. By using passively collected smartphone use and accelerometer data, the CT-HMM was able to identify rest-activity states and the effect of the hour of the day on the likelihood of being active or at rest, which we used to construct the ARM, and found it to be associated with a variety of demographic, sleep, and mood or depression variables. We validated the ARM relative to self-reported nightly sleep-wake cycles and found that the ARM was correlated to variability in sleep or wake times from Beibe surveys. It is important to note that additional follow-up studies are necessary to validate our ARM covariate relationships. Our primary analyses suggest that the ARM is a promising alternative to previous diary-based metrics, which are often used to assess treatment efficacy.



## Acknowledgments

This work was supported by grants from the National Institute of Mental Health (R01MH116884). The AE Foundation provided support for data collection (R01MH113550 and R01MH107703).

## Authors' Contributions

BR and IB made substantial contributions to the conception and design of the work; BR, IB, TS, CHX, and PG contributed substantially to the acquisition, analysis, or interpretation of data for the work. All the authors have read and approved the final version of the manuscript. All the authors have agreed to its publication.

## Conflicts of Interest

None declared.

## References

1. Torous J, Staples P, Onnela J. Realizing the potential of mobile mental health: new methods for new data in psychiatry. *Curr Psychiatry Rep* 2015 Aug 16;17(8):602 [FREE Full text] [doi: [10.1007/s11920-015-0602-0](https://doi.org/10.1007/s11920-015-0602-0)] [Medline: [26073363](https://pubmed.ncbi.nlm.nih.gov/26073363/)]
2. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology* 2018 Jul 22;43(8):1660-1666 [FREE Full text] [doi: [10.1038/s41386-018-0030-z](https://doi.org/10.1038/s41386-018-0030-z)] [Medline: [29511333](https://pubmed.ncbi.nlm.nih.gov/29511333/)]
3. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 2016 Jun 28;41(7):1691-1696 [FREE Full text] [doi: [10.1038/npp.2016.7](https://doi.org/10.1038/npp.2016.7)] [Medline: [26818126](https://pubmed.ncbi.nlm.nih.gov/26818126/)]
4. Lin Y, Wong B, Pan Y, Chiu Y, Lee Y. Validation of the mobile app-recorded circadian rhythm by a digital footprint. *JMIR Mhealth Uhealth* 2019 May 16;7(5):e13421 [FREE Full text] [doi: [10.2196/13421](https://doi.org/10.2196/13421)] [Medline: [31099340](https://pubmed.ncbi.nlm.nih.gov/31099340/)]
5. Cho C, Lee T, Kim M, In HP, Kim L, Lee H. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. *J Med Internet Res* 2019 Apr 17;21(4):e11029. [doi: [10.2196/11029](https://doi.org/10.2196/11029)]
6. Smith MT, McCrae CS, Cheung J, Martin JL, Harrod CG, Heald JL, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an american academy of sleep medicine clinical practice guideline. *J Clin Sleep Med* 2018 Jul 15;14(7):1231-1237 [FREE Full text] [doi: [10.5664/jcsm.7230](https://doi.org/10.5664/jcsm.7230)] [Medline: [29991437](https://pubmed.ncbi.nlm.nih.gov/29991437/)]
7. Seppälä J, De Vita I, Jämsä T, Miettunen J, Isohanni M, Rubinstein K, M-RESIST Group, et al. Mobile phone and wearable sensor-based mHealth approaches for psychiatric disorders and symptoms: systematic review. *JMIR Ment Health* 2019 Feb 20;6(2):e9819 [FREE Full text] [doi: [10.2196/mental.9819](https://doi.org/10.2196/mental.9819)] [Medline: [30785404](https://pubmed.ncbi.nlm.nih.gov/30785404/)]
8. Aubourg T, Demongeot J, Vuillerme N. Gaining insights into the estimation of the circadian rhythms of social activity in older adults from their telephone call activity with statistical learning: observational study. *J Med Internet Res* 2021 Jan 08;23(1):e22339 [FREE Full text] [doi: [10.2196/22339](https://doi.org/10.2196/22339)] [Medline: [33416502](https://pubmed.ncbi.nlm.nih.gov/33416502/)]
9. Klerman EB. Clinical aspects of human circadian rhythms. *J Biol Rhythms* 2005 Aug 29;20(4):375-386 [FREE Full text] [doi: [10.1177/0748730405278353](https://doi.org/10.1177/0748730405278353)] [Medline: [16077156](https://pubmed.ncbi.nlm.nih.gov/16077156/)]
10. Akerstedt T. Psychological and psychophysiological effects of shift work. *Scand J Work Environ Health* 1990;16 Suppl 1:67-73 [FREE Full text] [doi: [10.5271/sjweh.1819](https://doi.org/10.5271/sjweh.1819)] [Medline: [2189223](https://pubmed.ncbi.nlm.nih.gov/2189223/)]
11. Richardson GS, Malin HV. Circadian rhythm sleep disorders: pathophysiology and treatment. *J Clin Neurophysiol* 1996 Jan;13(1):17-31. [doi: [10.1097/00004691-199601000-00003](https://doi.org/10.1097/00004691-199601000-00003)] [Medline: [8988283](https://pubmed.ncbi.nlm.nih.gov/8988283/)]
12. Melo MC, Abreu RL, Linhares Neto VB, de Bruin PF, de Bruin VM. Chronotype and circadian rhythm in bipolar disorder: a systematic review. *Sleep Med Rev* 2017 Aug;34:46-58. [doi: [10.1016/j.smrv.2016.06.007](https://doi.org/10.1016/j.smrv.2016.06.007)] [Medline: [27524206](https://pubmed.ncbi.nlm.nih.gov/27524206/)]
13. Skarke C, Lahens NF, Rhoades SD, Campbell A, Bittinger K, Bailey A, et al. A pilot characterization of the human chronobiome. *Sci Rep* 2017 Dec 07;7(1):17141 [FREE Full text] [doi: [10.1038/s41598-017-17362-6](https://doi.org/10.1038/s41598-017-17362-6)] [Medline: [29215023](https://pubmed.ncbi.nlm.nih.gov/29215023/)]
14. Morris CJ, Aeschbach D, Scheer FA. Circadian system, sleep and endocrinology. *Mol Cell Endocrinol* 2012 Feb 05;349(1):91-104 [FREE Full text] [doi: [10.1016/j.mce.2011.09.003](https://doi.org/10.1016/j.mce.2011.09.003)] [Medline: [21939733](https://pubmed.ncbi.nlm.nih.gov/21939733/)]
15. Yang G, Chen L, Grant GR, Paschos G, Song W, Musiek ES, et al. Timing of expression of the core clock gene Bmal1 influences its effects on aging and survival. *Sci Transl Med* 2016 Feb 03;8(324):324ra16 [FREE Full text] [doi: [10.1126/scitranslmed.aad3305](https://doi.org/10.1126/scitranslmed.aad3305)] [Medline: [26843191](https://pubmed.ncbi.nlm.nih.gov/26843191/)]
16. Lange T, Dimitrov S, Born J. Effects of sleep and circadian rhythm on the human immune system. *Ann N Y Acad Sci* 2010 Apr;1193:48-59. [doi: [10.1111/j.1749-6632.2009.05300.x](https://doi.org/10.1111/j.1749-6632.2009.05300.x)] [Medline: [20398008](https://pubmed.ncbi.nlm.nih.gov/20398008/)]
17. Cederroth CR, Albrecht U, Bass J, Brown SA, Dyhrfjeld-Johnsen J, Gachon F, et al. Medicine in the fourth dimension. *Cell Metab* 2019 Aug 06;30(2):238-250 [FREE Full text] [doi: [10.1016/j.cmet.2019.06.019](https://doi.org/10.1016/j.cmet.2019.06.019)] [Medline: [31390550](https://pubmed.ncbi.nlm.nih.gov/31390550/)]
18. Tsanas A, Woodward E, Ehlers A. Objective characterization of activity, sleep, and circadian rhythm patterns using a wrist-worn actigraphy sensor: insights into posttraumatic stress disorder. *JMIR Mhealth Uhealth* 2020 Apr 20;8(4):e14306 [FREE Full text] [doi: [10.2196/14306](https://doi.org/10.2196/14306)] [Medline: [32310142](https://pubmed.ncbi.nlm.nih.gov/32310142/)]

19. Lyall LM, Wyse CA, Graham N, Ferguson A, Lyall DM, Cullen B, et al. Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. *Lancet Psychiatry* 2018 Jun;5(6):507-514. [doi: [10.1016/s2215-0366\(18\)30139-1](https://doi.org/10.1016/s2215-0366(18)30139-1)]
20. Monk TH, Kupfer DJ, Frank E, Ritenour AM. The social rhythm metric (SRM): measuring daily social rhythms over 12 weeks. *Psychiatry Res* 1991 Feb;36(2):195-207. [doi: [10.1016/0165-1781\(91\)90131-8](https://doi.org/10.1016/0165-1781(91)90131-8)]
21. Monk TH, Flaherty JF, Frank E, Hoskinson K, Kupfer DJ. The Social Rhythm Metric. An instrument to quantify the daily rhythms of life. *J Nerv Ment Dis* 1990 Feb;178(2):120-126. [doi: [10.1097/00005053-199002000-00007](https://doi.org/10.1097/00005053-199002000-00007)] [Medline: [2299336](https://pubmed.ncbi.nlm.nih.gov/2299336/)]
22. Alloy LB, Boland EM, Ng TH, Whitehouse WG, Abramson LY. Low social rhythm regularity predicts first onset of bipolar spectrum disorders among at-risk individuals with reward hypersensitivity. *J Abnorm Psychol* 2015 Nov;124(4):944-952 [FREE Full text] [doi: [10.1037/abn0000107](https://doi.org/10.1037/abn0000107)] [Medline: [26595474](https://pubmed.ncbi.nlm.nih.gov/26595474/)]
23. Shen GH, Alloy LB, Abramson LY, Sylvia LG. Social rhythm regularity and the onset of affective episodes in bipolar spectrum individuals. *Bipolar Disord* 2008 Jun;10(4):520-529 [FREE Full text] [doi: [10.1111/j.1399-5618.2008.00583.x](https://doi.org/10.1111/j.1399-5618.2008.00583.x)] [Medline: [18452448](https://pubmed.ncbi.nlm.nih.gov/18452448/)]
24. Levenson JC, Wallace ML, Anderson BP, Kupfer DJ, Frank E. Social rhythm disrupting events increase the risk of recurrence among individuals with bipolar disorder. *Bipolar Disord* 2015 Dec 28;17(8):869-879 [FREE Full text] [doi: [10.1111/bdi.12351](https://doi.org/10.1111/bdi.12351)] [Medline: [26614534](https://pubmed.ncbi.nlm.nih.gov/26614534/)]
25. Shear MK, Randall J, Monk TH, Ritenour A, Tu X, Frank E, et al. Social rhythm in anxiety disorder patients. *Anxiety* 1994 Nov 07;1(2):90-95. [doi: [10.1002/anxi.3070010208](https://doi.org/10.1002/anxi.3070010208)] [Medline: [9160553](https://pubmed.ncbi.nlm.nih.gov/9160553/)]
26. Szuba MP, Yager A, Guze BH, Allen EM, Baxter LR. Disruption of social circadian rhythms in major depression: a preliminary report. *Psychiatry Res* 1992 Jun;42(3):221-230. [doi: [10.1016/0165-1781\(92\)90114-j](https://doi.org/10.1016/0165-1781(92)90114-j)]
27. Haynes PL, Gengler D, Kelly M. Social rhythm therapies for mood disorders: an update. *Curr Psychiatry Rep* 2016 Aug 24;18(8):75 [FREE Full text] [doi: [10.1007/s11920-016-0712-3](https://doi.org/10.1007/s11920-016-0712-3)] [Medline: [27338753](https://pubmed.ncbi.nlm.nih.gov/27338753/)]
28. Hallonquist J, Goldberg M, Brandes J. Affective disorders and circadian rhythms. *Can J Psychiatry* 1986 Apr 01;31(3):259-272. [doi: [10.1177/070674378603100315](https://doi.org/10.1177/070674378603100315)] [Medline: [3518905](https://pubmed.ncbi.nlm.nih.gov/3518905/)]
29. Boland EM, Goldschmied JR, Kelly MR, Perkins S, Gehrman PR, Haynes PL. Social rhythm regularity moderates the relationship between sleep disruption and depressive symptoms in veterans with post-traumatic stress disorder and major depressive disorder. *Chronobiol Int* 2019 Oct 01;36(10):1429-1438 [FREE Full text] [doi: [10.1080/07420528.2019.1644344](https://doi.org/10.1080/07420528.2019.1644344)] [Medline: [31368369](https://pubmed.ncbi.nlm.nih.gov/31368369/)]
30. Hasler BP, Smith LJ, Cousins JC, Bootzin RR. Circadian rhythms, sleep, and substance abuse. *Sleep Med Rev* 2012 Feb;16(1):67-81 [FREE Full text] [doi: [10.1016/j.smrv.2011.03.004](https://doi.org/10.1016/j.smrv.2011.03.004)] [Medline: [21620743](https://pubmed.ncbi.nlm.nih.gov/21620743/)]
31. Conroy DA, Arnedt JT. Sleep and substance use disorders: an update. *Curr Psychiatry Rep* 2014 Oct 19;16(10):487. [doi: [10.1007/s11920-014-0487-3](https://doi.org/10.1007/s11920-014-0487-3)] [Medline: [25135784](https://pubmed.ncbi.nlm.nih.gov/25135784/)]
32. Takayanagi Y, Spira AP, Roth KB, Gallo JJ, Eaton WW, Mojtabai R. Accuracy of reports of lifetime mental and physical disorders: results from the Baltimore Epidemiological Catchment Area study. *JAMA Psychiatry* 2014 Mar 01;71(3):273-280 [FREE Full text] [doi: [10.1001/jamapsychiatry.2013.3579](https://doi.org/10.1001/jamapsychiatry.2013.3579)] [Medline: [24402003](https://pubmed.ncbi.nlm.nih.gov/24402003/)]
33. Trifan A, Oliveira M, Oliveira JL. Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. *JMIR Mhealth Uhealth* 2019 Aug 23;7(8):e12649 [FREE Full text] [doi: [10.2196/12649](https://doi.org/10.2196/12649)] [Medline: [31444874](https://pubmed.ncbi.nlm.nih.gov/31444874/)]
34. Lawson S, Jamison-Powell S, Garbett A, Linehan C, Kucharczyk E, Verbaan S, et al. Validating a mobile phone application for the everyday, unobtrusive, objective measurement of sleep. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; Apr 27-May 2, 2013; Paris France. [doi: [10.1145/2470654.2481345](https://doi.org/10.1145/2470654.2481345)]
35. Chen Z, Lin M, Chen F, Lane N, Cardone G, Wang R, et al. Unobtrusive sleep monitoring using smartphones. In: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops. 2013 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops; May 05-08, 2013; Venice, Italy.
36. Hao T, Xing G, Zhou G. iSleep: unobtrusive sleep quality monitoring using smartphones. In: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. 2013 Presented at: SenSys '13: The 11th ACM Conference on Embedded Network Sensor Systems; Nov 11 - 15, 2013; Roma Italy.
37. Melo MC, Garcia RF, Linhares Neto VB, Sá MB, de Mesquita LM, de Araújo CF, et al. Sleep and circadian alterations in people at risk for bipolar disorder: a systematic review. *J Psychiatr Res* 2016 Dec;83:211-219. [doi: [10.1016/j.jpsychires.2016.09.005](https://doi.org/10.1016/j.jpsychires.2016.09.005)] [Medline: [27661417](https://pubmed.ncbi.nlm.nih.gov/27661417/)]
38. Van Reeth O, Weibel L, Spiegel K, Leproult R, Dugovic C, Maccari S. PHYSIOLOGY OF SLEEP (REVIEW)–Interactions between stress and sleep: from basic research to clinical situations. *Sleep Med Rev* 2000 Apr;4(2):201-219 [FREE Full text] [doi: [10.1053/smrv.1999.0097](https://doi.org/10.1053/smrv.1999.0097)]
39. Sheikh M, Qassem M, Kyriacou PA. Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. *Front Digit Health* 2021 Apr 7;3:662811 [FREE Full text] [doi: [10.3389/fdgh.2021.662811](https://doi.org/10.3389/fdgh.2021.662811)] [Medline: [34713137](https://pubmed.ncbi.nlm.nih.gov/34713137/)]

40. Thompson S, Fergus P, Chalmers C, Reilly D. Detection of obstructive sleep Apnoea using features extracted from segmented time-series ECG signals using a one dimensional convolutional neural network. In: Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN). 2020 Presented at: 2020 International Joint Conference on Neural Networks (IJCNN); Jul 19-24, 2020; Glasgow, UK.
41. Palotti J, Mall R, Aupetit M, Rueschman M, Singh M, Sathyanarayana A, et al. Benchmark on a large cohort for sleep-wake classification with machine learning techniques. NPJ Digit Med 2019 Jun 7;2(1):50 [FREE Full text] [doi: [10.1038/s41746-019-0126-9](https://doi.org/10.1038/s41746-019-0126-9)] [Medline: [31304396](https://pubmed.ncbi.nlm.nih.gov/31304396/)]
42. Chen Z, Wu M, Gao K, Wu J, Ding J, Zeng Z, et al. A novel ensemble deep learning approach for sleep-wake detection using heart rate variability and acceleration. IEEE Trans Emerg Top Comput Intell 2021 Oct;5(5):803-812 [FREE Full text] [doi: [10.1109/tetci.2020.2996943](https://doi.org/10.1109/tetci.2020.2996943)]
43. Koslovsky MD, Hébert ET, Businelle MS, Vannucci M. A Bayesian time-varying effect model for behavioral mHealth data. Ann Appl Stat 2020 Dec 1;14(4):1878-1902 [FREE Full text] [doi: [10.1214/20-aos1402](https://doi.org/10.1214/20-aos1402)] [Medline: [35386276](https://pubmed.ncbi.nlm.nih.gov/35386276/)]
44. Dempsey W, Moreno A, Scott C, Dennis M, Gustafson D, Murphy S, et al. iSurvive: an interpretable, event-time prediction model for mHealth. Proc Mach Learn Res 2017;70:970-979 [FREE Full text] [Medline: [30906932](https://pubmed.ncbi.nlm.nih.gov/30906932/)]
45. Huang Q, Cohen D, Komarzynski S, Li X, Innominato P, Lévi F, et al. Hidden Markov models for monitoring circadian rhythmicity in telemetric activity data. J R Soc Interface 2018 Feb 07;15(139):20170885 [FREE Full text] [doi: [10.1098/rsif.2017.0885](https://doi.org/10.1098/rsif.2017.0885)] [Medline: [29436510](https://pubmed.ncbi.nlm.nih.gov/29436510/)]
46. Scherer EA, Ben-Zeev D, Li Z, Kane JM. Analyzing mHealth engagement: joint models for intensively collected user engagement data. JMIR Mhealth Uhealth 2017 Jan 12;5(1):e1 [FREE Full text] [doi: [10.2196/mhealth.6474](https://doi.org/10.2196/mhealth.6474)] [Medline: [28082257](https://pubmed.ncbi.nlm.nih.gov/28082257/)]
47. van Kasteren T, Englebienne G, Kröse B. Activity recognition using semi-Markov models on real world smart home datasets. J Ambient Intell Smart Environ 2010;311-325. [doi: [10.3233/AIS-2010-0070](https://doi.org/10.3233/AIS-2010-0070)]
48. Kim E, Helal S, Cook D. Human activity recognition and pattern discovery. IEEE Pervasive Comput 2010;9(1):48 [FREE Full text] [doi: [10.1109/MPRV.2010.7](https://doi.org/10.1109/MPRV.2010.7)] [Medline: [21258659](https://pubmed.ncbi.nlm.nih.gov/21258659/)]
49. Natarajan P, Nevatia R. Coupled hidden semi Markov models for activity recognition. In: Proceedings of the 2007 IEEE Workshop on Motion and Video Computing (WMVC'07). 2007 Presented at: 2007 IEEE Workshop on Motion and Video Computing (WMVC'07); Feb 23-24, 2007; Austin, TX, USA.
50. Enshaeifar S, Zoha A, Markides A, Skillman S, Acton ST, Elsaleh T, et al. Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques. PLoS One 2018 May 3;13(5):e0195605 [FREE Full text] [doi: [10.1371/journal.pone.0195605](https://doi.org/10.1371/journal.pone.0195605)] [Medline: [29723236](https://pubmed.ncbi.nlm.nih.gov/29723236/)]
51. Alshurafa N, Xu W, Liu JJ, Huang M, Mortazavi B, Roberts CK, et al. Designing a robust activity recognition framework for health and exergaming using wearable sensors. IEEE J Biomed Health Inform 2014 Sep;18(5):1636-1646. [doi: [10.1109/jbhi.2013.2287504](https://doi.org/10.1109/jbhi.2013.2287504)]
52. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y. Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: a systematic review. J Biomed Inform 2018 Nov;87:138-153 [FREE Full text] [doi: [10.1016/j.jbi.2018.09.002](https://doi.org/10.1016/j.jbi.2018.09.002)] [Medline: [30267895](https://pubmed.ncbi.nlm.nih.gov/30267895/)]
53. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz J. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In: Ambient Assisted Living and Home Care. Berlin, Heidelberg: Springer; 2012.
54. Costa MJ, Finkenstädt B, Roche V, Lévi F, Gould PD, Foreman J, et al. Inference on periodicity of circadian time series. Biostatistics 2013 Sep 06;14(4):792-806 [FREE Full text] [doi: [10.1093/biostatistics/kxt020](https://doi.org/10.1093/biostatistics/kxt020)] [Medline: [23743206](https://pubmed.ncbi.nlm.nih.gov/23743206/)]
55. Wang J, Xian H, Licis A, Deych E, Ding J, McLeland J, et al. Measuring the impact of apnea and obesity on circadian activity patterns using functional linear modeling of actigraphy data. J Circadian Rhythms 2011 Oct 13;9(1):11 [FREE Full text] [doi: [10.1186/1740-3391-9-11](https://doi.org/10.1186/1740-3391-9-11)] [Medline: [21995417](https://pubmed.ncbi.nlm.nih.gov/21995417/)]
56. Ruf T. The Lomb-Scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series. Biological Rhythm Res 2010 Aug 09;30(2):178-201 [FREE Full text] [doi: [10.1076/brhm.30.2.178.1422](https://doi.org/10.1076/brhm.30.2.178.1422)]
57. Langrock R, Swihart BJ, Caffo BS, Punjabi NM, Crainiceanu CM. Combining hidden Markov models for comparing the dynamics of multiple sleep electroencephalograms. Stat Med 2013 Aug 30;32(19):3342-3356 [FREE Full text] [doi: [10.1002/sim.5747](https://doi.org/10.1002/sim.5747)] [Medline: [23348835](https://pubmed.ncbi.nlm.nih.gov/23348835/)]
58. Pan S, Kuo C, Zeng J, Liang S. A transition-constrained discrete hidden Markov model for automatic sleep staging. BioMed Eng OnLine 2012 Aug 21;11(1):52. [doi: [10.1186/1475-925x-11-52](https://doi.org/10.1186/1475-925x-11-52)]
59. Stoner O, Economou T. An advanced hidden Markov model for hourly rainfall time series. Computational Stat Data Analysis 2020 Dec;152:107045 [FREE Full text] [doi: [10.1016/j.csda.2020.107045](https://doi.org/10.1016/j.csda.2020.107045)]
60. Holsclaw T, Greene AM, Robertson AW, Smyth P. Bayesian nonhomogeneous Markov models via Pólya-Gamma data augmentation with applications to rainfall modeling. Ann Appl Stat 2017 Mar 1;11(1):393-426 [FREE Full text] [doi: [10.1214/16-aos1009](https://doi.org/10.1214/16-aos1009)]
61. Altman RM. Mixed hidden Markov models. J Am Stat Assoc 2007 Mar;102(477):201-210 [FREE Full text] [doi: [10.1198/01621450600001086](https://doi.org/10.1198/01621450600001086)]

62. Xia C, Barnett I, Tapera T, Adebimpe A, Baker JT, Bassett DS, et al. Mobile footprinting: linking individual distinctiveness in mobility patterns to mood, sleep, and brain functional connectivity. *Neuropsychopharmacology* 2022 Aug;47(9):1662-1671 [[FREE Full text](#)] [doi: [10.1038/s41386-022-01351-z](https://doi.org/10.1038/s41386-022-01351-z)] [Medline: [35660803](https://pubmed.ncbi.nlm.nih.gov/35660803/)]
63. Calkins ME, Moore TM, Satterthwaite TD, Wolf DH, Turetsky BI, Roalf DR, et al. Persistence of psychosis spectrum symptoms in the Philadelphia Neurodevelopmental Cohort: a prospective two-year follow-up. *World Psychiatry* 2017 Mar 26;16(1):62-76 [[FREE Full text](#)] [doi: [10.1002/wps.20386](https://doi.org/10.1002/wps.20386)] [Medline: [28127907](https://pubmed.ncbi.nlm.nih.gov/28127907/)]
64. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Ment Health* 2016 May 05;3(2):e16 [[FREE Full text](#)] [doi: [10.2196/mental.5165](https://doi.org/10.2196/mental.5165)] [Medline: [27150677](https://pubmed.ncbi.nlm.nih.gov/27150677/)]
65. 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](https://doi.org/10.1016/0165-1781(89)90047-4)]
66. Beck AT. An inventory for measuring depression. *Arch Gen Psychiatry* 1961 Jun 01;4(6):561-571. [doi: [10.1001/archpsyc.1961.01710120031004](https://doi.org/10.1001/archpsyc.1961.01710120031004)] [Medline: [13688369](https://pubmed.ncbi.nlm.nih.gov/13688369/)]
67. Bulla J, Bulla I, Nenadić O. hsmm — An R package for analyzing hidden semi-Markov models. *Computational Stat Data Analysis* 2010 Mar;54(3):611-619. [doi: [10.1016/j.csda.2008.08.025](https://doi.org/10.1016/j.csda.2008.08.025)]
68. Fossion R, Rivera AL, Toledo-Roy JC, Ellis J, Angelova M. Multiscale adaptive analysis of circadian rhythms and intradaily variability: application to actigraphy time series in acute insomnia subjects. *PLoS ONE* 2017 Jul 28;12(7):e0181762. [doi: [10.1371/journal.pone.0181762](https://doi.org/10.1371/journal.pone.0181762)]
69. Bureau A, Shiboski S, Hughes JP. Applications of continuous time hidden Markov models to the study of misclassified disease outcomes. *Stat Med* 2003 Mar 15;22(3):441-462 [[FREE Full text](#)] [doi: [10.1002/sim.1270](https://doi.org/10.1002/sim.1270)] [Medline: [12529874](https://pubmed.ncbi.nlm.nih.gov/12529874/)]
70. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the algorithm. *J Royal Stat Soc Series B (Methodological)* 2018 Dec 05;39(1):1-22. [doi: [10.1111/j.2517-6161.1977.tb01600.x](https://doi.org/10.1111/j.2517-6161.1977.tb01600.x)]
71. Liu Y, Li S, Li F, Song L, Rehg J. Efficient learning of continuous-time hidden Markov models for disease progression. *Adv Neural Inf Process Syst* 2015;28:3599-3607 [[FREE Full text](#)] [Medline: [27019571](https://pubmed.ncbi.nlm.nih.gov/27019571/)]
72. Bulla J, Berzel A. Computational issues in parameter estimation for stationary hidden Markov models. *Computational Stat* 2007 Jul 13;23(1):1-18 [[FREE Full text](#)] [doi: [10.1007/s00180-007-0063-y](https://doi.org/10.1007/s00180-007-0063-y)]
73. Monk T, Petrie S, Hayes A, Kupfer D. Regularity of daily life in relation to personality, age, gender, sleep quality and circadian rhythms. *J Sleep Res* 1994 Dec;3(4):196-205 [[FREE Full text](#)] [doi: [10.1111/j.1365-2869.1994.tb00132.x](https://doi.org/10.1111/j.1365-2869.1994.tb00132.x)] [Medline: [10607126](https://pubmed.ncbi.nlm.nih.gov/10607126/)]
74. Monk TH, Reynolds CF, Kupfer DJ, Hoch CC, Carrier J, Houck PR. Differences over the life span in daily life-style regularity. *Chronobiol Int* 1997 May 07;14(3):295-306. [doi: [10.3109/07420529709001421](https://doi.org/10.3109/07420529709001421)] [Medline: [9167890](https://pubmed.ncbi.nlm.nih.gov/9167890/)]
75. Monk TH, Reynolds CF, Buysse DJ, DeGrazia JM, Kupfer DJ. The relationship between lifestyle regularity and subjective sleep quality. *Chronobiol Int* 2003 Jan 07;20(1):97-107. [doi: [10.1081/cbi-120017812](https://doi.org/10.1081/cbi-120017812)] [Medline: [12638693](https://pubmed.ncbi.nlm.nih.gov/12638693/)]
76. Sylvia LG, Alloy LB, Hafner JA, Gauger MC, Verdon K, Abramson LY. Life events and social rhythms in bipolar spectrum disorders: a prospective study. *Behav Ther* 2009 Jun;40(2):131-141 [[FREE Full text](#)] [doi: [10.1016/j.beth.2008.04.003](https://doi.org/10.1016/j.beth.2008.04.003)] [Medline: [19433144](https://pubmed.ncbi.nlm.nih.gov/19433144/)]
77. Monk TH, Frank E, Potts JM, Kupfer DJ. A simple way to measure daily lifestyle regularity. *J Sleep Res* 2002 Sep;11(3):183-190 [[FREE Full text](#)] [doi: [10.1046/j.1365-2869.2002.00300.x](https://doi.org/10.1046/j.1365-2869.2002.00300.x)] [Medline: [12220313](https://pubmed.ncbi.nlm.nih.gov/12220313/)]
78. Rieger A, Gaines A, Barnett I, Baldassano CF, Connolly Gibbons MB, Crits-Christoph P. Psychiatry outpatients' willingness to share social media posts and smartphone data for research and clinical purposes: survey study. *JMIR Form Res* 2019 Aug 29;3(3):e14329 [[FREE Full text](#)] [doi: [10.2196/14329](https://doi.org/10.2196/14329)] [Medline: [31493326](https://pubmed.ncbi.nlm.nih.gov/31493326/)]

## Abbreviations

**ARM:** Activity Rhythm Metric  
**BDI:** Beck Depression Inventory  
**CT-HMM:** continuous-time hidden Markov model  
**EM:** expectation-maximization  
**EMA:** ecological momentary assessment  
**HMM:** hidden Markov model  
**LRT:** likelihood ratio test  
**mHealth:** mobile health  
**PH:** proportional hazard  
**PSQI:** Pittsburgh Sleep Quality Index  
**SRM:** Social Rhythm Metric

*Edited by A Mavragani; submitted 28.09.21; peer-reviewed by P Haynes, D Adler; comments to author 09.11.21; revised version received 18.01.22; accepted 19.07.22; published 14.09.22.*

*Please cite as:*

*Ren B, Xia CH, Gehrman P, Barnett I, Satterthwaite T*

*Measuring Daily Activity Rhythms in Young Adults at Risk of Affective Instability Using Passively Collected Smartphone Data: Observational Study*

*JMIR Form Res 2022;6(9):e33890*

*URL: <https://formative.jmir.org/2022/9/e33890>*

*doi: [10.2196/33890](https://doi.org/10.2196/33890)*

*PMID: [36103225](https://pubmed.ncbi.nlm.nih.gov/36103225/)*

©Benny Ren, Cedric Huchuan Xia, Philip Gehrman, Ian Barnett, Theodore Satterthwaite. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 14.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Measurement of Heart Rate Using the Withings ScanWatch Device During Free-living Activities: Validation Study

Oonagh M Giggins<sup>1</sup>, BSc, MSc, PhD; Julie Doyle<sup>1</sup>, BSc, PhD; Suzanne Smith<sup>1</sup>, BSocSc, MSc; Daniel R Crabtree<sup>2</sup>, BSc, MSc, PhD; Matthew Fraser<sup>2</sup>, BSc, MSc, PhD

<sup>1</sup>NetwellCASALA, Dundalk Institute of Technology, Dundalk, Ireland

<sup>2</sup>Division of Biomedical Sciences, University of the Highlands and Islands, Inverness, United Kingdom

**Corresponding Author:**

Oonagh M Giggins, BSc, MSc, PhD

NetwellCASALA

Dundalk Institute of Technology

Dublin Road

Dundalk, A91 K584

Ireland

Phone: 353 429370200 ext 2114

Email: [oonagh.giggins@dkit.ie](mailto:oonagh.giggins@dkit.ie)

## Abstract

**Background:** Wrist-worn devices that incorporate photoplethysmography (PPG) sensing represent an exciting means of measuring heart rate (HR). A number of studies have evaluated the accuracy of HR measurements produced by these devices in controlled laboratory environments. However, it is also important to establish the accuracy of measurements produced by these devices outside the laboratory, in real-world, consumer use conditions.

**Objective:** This study sought to examine the accuracy of HR measurements produced by the Withings ScanWatch during free-living activities.

**Methods:** A sample of convenience of 7 participants volunteered (3 male and 4 female; mean age 64, SD 10 years; mean height 164, SD 4 cm; mean weight 77, SD 16 kg) to take part in this real-world validation study. Participants were instructed to wear the ScanWatch for a 12-hour period on their nondominant wrist as they went about their day-to-day activities. A Polar H10 heart rate sensor was used as the criterion measure of HR. Participants used a study diary to document activities undertaken during the 12-hour study period. These activities were classified according to the 11 following domains: desk work, eat or drink, exercise, gardening, household activities, self-care, shopping, sitting, sleep, travel, and walking. Validity was assessed using the Bland-Altman analysis, concordance correlation coefficient (CCC), and mean absolute percentage error (MAPE).

**Results:** Across all activity domains, the ScanWatch measured HR with MAPE values <10%, except for the shopping activity domain (MAPE=10.8%). The activity domains that were more sedentary in nature (eg, desk work, eat or drink, and sitting) produced the most accurate HR measurements with a small mean bias and MAPE values <5%. Moderate to strong correlations (CCC=0.526-0.783) were observed between devices for all activity domains, except during the walking activity domain, which demonstrated a weak correlation (CCC=0.164) between devices.

**Conclusions:** The results of this study show that the ScanWatch measures HR with a degree of accuracy that is acceptable for general consumer use; however, it would not be suitable in circumstances where more accurate measurements of HR are required, such as in health care or in clinical trials. Overall, the ScanWatch was less accurate at measuring HR during ambulatory activities (eg, walking, gardening, and household activities) compared to more sedentary activities (eg, desk work, eat or drink, and sitting). Further larger-scale studies examining this device in different populations and during different activities are required.

(*JMIR Form Res* 2022;6(9):e34280) doi:[10.2196/34280](https://doi.org/10.2196/34280)

**KEYWORDS**

heart rate; photoplethysmography; PPG; wearable electronic device; wrist-worn device; validation study; heart; activity; physical activity; free-living activity

## Introduction

The consumer wearable device industry has rapidly grown in recent years, with a wide range of devices available for monitoring and tracking physical activity [1]. In addition to providing physical activity metrics, such as step counts, distance covered, and energy expenditure, many devices now also incorporate an optical sensor that estimates the wearer's heart rate (HR). These optical sensors estimate HR by means of a technique called photoplethysmography (PPG). This noninvasive optical technique detects blood volume changes in the microvascular bed of tissue beneath the skin [2]. Traditionally, wearable devices for monitoring HR have used an electrode-based chest strap to detect the electrical signals from the heart (eg, Polar). These devices have been extensively validated [3,4], and although much more convenient than an electrocardiogram (ECG) Holter monitor, these chest straps are not always feasible, desirable, or comfortable, particularly for prolonged use [5]. Therefore, wrist-worn measurement of HR via PPG represents a more unobtrusive and comfortable means of monitoring HR.

The potential to inform improved health care delivery is offered by wrist-worn PPG devices with screening or diagnostic, therapeutic monitoring, and self-management applications [6-9]. As the popularity and availability of PPG-enabled devices increases, it becomes more important that devices produce accurate measurements of HR, particularly if used in health care settings, where inaccurate readings could have detrimental consequences. Many studies have already been conducted examining the accuracy of HR measurements derived from wrist-worn devices equipped with PPG [10-13]. However, given the pace of technological development, these studies have largely been conducted on devices that are no longer available or that have been replaced by newer-generation models. Continual testing and validation of devices are, therefore, necessary to keep abreast of the advancements being made in this field. It is recommended that these devices should be evaluated in settings appropriate for their intended use [14]. Testing of devices in a controlled laboratory environment is initially required to establish validity of measurements. However, in laboratory evaluations, the range of activity types and intensities are often limited. Testing devices outside of the laboratory in more naturalistic environments during activities of daily living, where movements are more variable and sporadic, supports an understanding of how such devices are likely to perform in real-world context.

We have conducted this validation study to assess the accuracy of HR measurements produced by the Withings ScanWatch under free-living conditions. The Withings ScanWatch has an embedded PPG sensor for measuring HR as well as oxygen saturation, a triaxial accelerometer for monitoring activity, and electrodes for electrogram recording. To our knowledge, no study to date has examined HR measurements obtained from the ScanWatch, in the laboratory or in real-world conditions, as intended for general consumer use. Testing of the Withings ScanWatch was undertaken as part of evaluations of devices for use in trials conducted in our laboratory.

## Methods

### Study Design

A cross-sectional validation study was conducted to examine the accuracy of HR measurements obtained from the Withings ScanWatch. Data were collected from volunteer participants during free-living activities.

### Ethics Approval

The study protocol was approved by the Health and Science Ethics Committee in Dundalk Institute of Technology, and all procedures were conducted in accordance with the Declaration of Helsinki 1974 and its later amendments.

### Data Collection

A convenience sample of adult volunteers participated in the study. Participants were members of the NetwellCASALA Living Lab Panel in Dundalk Institute of Technology and were invited to participate via email and a newsletter sent to all panel participants (n=18). Informed consent was obtained from members upon enrolment to the Living Lab Panel, to take part in device testing research activities. Participants were provided with a study information leaflet and notified that they were not obliged to take part in this investigation and were free to withdraw from it at any point.

All panel members were eligible to participate. The only exclusion criteria that applied were the following: any cardiometabolic conditions, pacemakers, or implanted electrical devices that could interfere with HR measurements, as well as undertaking any activities or work that could interfere with the completion of the study diary. Participants were screened by telephone to ensure suitability to take part in this investigation.

This study took place between April and June 2021, during level 5 of COVID-19 restrictions in Ireland. This prevented researcher visits to participants' homes for technology deployment and data collection. Therefore, demographic data and Fitzpatrick skin tone measurements [15] were obtained verbally from participants via telephone. The study materials and required devices were delivered to participants' homes by a member of the research team.

Participants were instructed via a Zoom videoconference call (Zoom Video Communications, Inc) on how to position the devices on their body for the study. The Withings ScanWatch was worn on participants' nondominant wrist. A Polar H10 HR sensor (Polar Electro) was used as the criterion measure of HR. This type of monitor has been shown to be highly accurate in measuring HR [16]. The Polar chest strap was dampened and placed following Polar's guidelines. Participants were also provided with written and pictorial instructions in the form of a study manual ([Multimedia Appendix 1](#)).

The Polar H10 was paired with the Polar Beat app on an Android smartphone for data acquisition, while the ScanWatch was paired with the Withings Health Mate app. Recordings commenced at 9 AM and continued for a 12-hour period until 9 PM. Participants were not asked to interact in any way with the watch or the Polar device during the 12-hour study period. Participants used a paper-based study diary to document all

activities undertaken as well as the start and end times associated with each. Activities undertaken during the 12-hour period were classified according to the 11 following domains: (1) desk work—any activity sitting at a desk (eg, typing, emailing, writing, surfing the internet, phone calls, and video conferencing); (2) eat or drink (ie, sitting down for meals or snacks and drinks); (3) exercise—a defined and structured bout of moderate to intense physical activity (ie, jogging, running, and cycling); (4) gardening (eg, general garden maintenance, raking, weeding, and mowing the lawn); (5) household activities (ie, laundry, ironing, general home maintenance, vacuum cleaning, washing floors, washing windows, food preparation, cooking, setting table, and washing or putting away dishes); (6) self-care (eg, wash, dress, and care for self); (7) shopping (ie, purchasing goods or consuming other services in supermarkets and shops); (8) sitting—any period of quiet sitting activity, including watching TV, reading, and listening to music; (9) sleep (ie, any sleep and naps in bed, chairs, or reclined position); (10) travel (ie, travel by car, bus, and train); (11) walking—any purposeful walking activity of more than 1 minute duration (eg, walking the dog, walking to work, and walking to shop). The transitions between activities (eg, from sitting to standing or from standing to sitting) were not removed from the data, as per previous work [17].

### Data Processing

Following data collection, a member of the research team collected the devices and the completed study diary from participants. The ScanWatch was synchronized with the Withings Health Mate app, and the Polar Beat data was synchronized with the Polar Flow web application. Data were downloaded from both the Polar Flow and Health Mate web applications in CSV file format for analysis. The Polar H10 collected second-by-second HR data, whereas the ScanWatch measured HR approximately every 10 minutes. When in workout mode, the ScanWatch has a higher sampling frequency; however, in order to simulate real-world use, the workout mode was not used in this study. The Polar H10 data were plotted linearly to visually examine data quality and check for errors. Data from both devices were time-aligned and split according to their corresponding activity classification. Data alignment was performed by visually inspecting the data files using Microsoft Excel (Version 16) and manually matching the corresponding data points according to their timestamp. The timestamps used were provided via data export from the internal clock of each device, which was paired and synchronized with the same Android smartphone. The Polar H10 data were averaged in 60-second epochs, and comparisons between the ScanWatch and Polar H10 data were performed for each matched timestamp.

### Statistical Analysis

Descriptive statistics of the mean and SD were used to summarize all data collected. ScanWatch accuracy was assessed by calculating the difference between the ScanWatch-measured and the Polar H10-measured HR in beats per minute (BPM) for each activity domain. Mean absolute percentage error (MAPE) values were calculated as  $(|HR_{polar} - HR_{ScanWatch}| / HR_{polar} \times 100)$ . To assess the degree of agreement between

the two devices, Lin's Concordance Correlation Coefficient (CCC) and the 95% CI were calculated [14].

The strength of agreement was interpreted as follows: weak ( $CCC < 0.5$ ), moderate ( $CCC = 0.5-0.7$ ), and strong ( $CCC > 0.7$ ) [17]. The Bland-Altman method was also used to express agreement between the Polar H10-measured and ScanWatch-measured HR, with limits of agreement (LoA) calculated as 1.96 times the SD of the mean difference [18].

## Results

A total of 6 Living Lab Panel members volunteered to take part in this study, and an additional 3 participants were recruited from within the research team. Due to a data syncing issue, Polar H10 data were missing for 2 participants, and therefore, data from 7 participants (3 male and 4 female; mean age 64, SD 10 years; mean height 164, SD 4 cm; mean weight 77, SD 16 kg; Fitzpatrick skin tone:  $n=2$  for type II,  $n=3$  for type III, and  $n=1$  for type V) were analyzed. A number of ScanWatch data points were missing for each participant over the 12-hour study period, and therefore, the expected number of 504 data points was not achieved. A total of 422 matched timestamped data points from all participants were available for analysis. Table 1 outlines the number of data points analyzed for each activity domain. There were less than 10 matched data points available for analysis in the domains of exercise, self-care, and sleep. Data for these activity domains were therefore only descriptively analyzed. Correlation coefficients, mean bias, 95% LoA, and MAPE values across 12 hours and for each activity domain are also indicated in Table 1.

Combined data across the 12 hours showed that the ScanWatch marginally underestimated HR with a mean bias of 1.5 (SD 8.4) BPM (95% LoA 18.0-14.9). There was a strong correlation between the ScanWatch-measured HR and the criterion device's measured HR ( $CCC=0.796$ , 95% CI 0.759-0.828). The MAPE for the ScanWatch across the 12 hours was 5.1%. The distribution of the error is presented in the Bland-Altman plot in Figure 1.

Figure 2 presents the Bland-Altman plots for all other activity domains. The activity domains that involved sitting, namely desk work, eat or drink, and sitting all produced a small mean bias and MAPE values  $< 5\%$  (desk work mean bias 0.8 (SD 1.1) BPM, MAPE=4.2%; eat or drink mean bias 1.6 (SD 5.9) BPM, MAPE=3.9%; and sitting mean bias 0.9 (SD 5.6) BPM, MAPE=3.8%). Strong correlations were observed for the desk work ( $CCC=0.827$ , 95% CI 0.748-0.883), eat or drink ( $CCC=0.796$ , 95% CI 0.678-0.875), and sitting ( $CCC=0.865$ , 95% CI 0.783-0.918) activity domains. All other activity domains produced MAPE values  $< 10\%$ , except for shopping (Table 1). The ScanWatch exhibited a weak correlation with the criterion measure of HR during the walking activity domain, underestimating HR with a mean bias of 6.6 (SD 15.0) BPM. During the activity domains gardening, household activities, and shopping, the ScanWatch had a tendency to underestimate HR, with fair to moderate correlations observed. The ScanWatch overestimated HR during the travel domain with a mean bias of  $-1.3$  (SD 9.5) BPM and MAPE of 6.5%.



**Table 1.** Validity of measuring heart rate using the Withings ScanWatch.

Activity domain	Data points <sup>a</sup> , n	Polar H10, mean (SD)	Withings ScanWatch				
			Mean (SD)	Mean bias (SD)	95% LoA <sup>b</sup> (upper, lower)	MAPE <sup>c</sup> (%)	CCC <sup>d</sup> (95% CI)
12-hours	422	79.2 (12.9)	77.7 (13.7)	1.5 (8.4)	18.0, -14.9	5.1	0.796 (0.759-0.828)
Desk work	87	79.5 (12.1)	78.8 (13.1)	0.8 (1.1)	2.9, -1.4	4.2	0.827 (0.748-0.883)
Eat or drink	56	77.7 (9.5)	76.1 (9.5)	1.6 (5.9)	13.1, -9.9	3.9	0.796 (0.678-0.875)
Exercise	3	96.3 (30.0)	94.3 (28.3)	— <sup>e</sup>	—	—	—
Gardening	48	82.3 (14.0)	79.9 (15.5)	2.4 (10.0)	21.9, -17.1	6.3	0.762 (0.617-0.858)
Household activities	100	81.7 (9.9)	79.9 (12.0)	1.8 (8.4)	18.3, -14.7	5	0.696 (0.586-0.782)
Self-care	5	88.6 (11.2)	91.6 (11.2)	—	—	—	—
Shopping	16	72.8 (12.2)	69.8 (14.7)	3.0 (12.1)	26.8, -20.8	10.8	0.582 (0.165-0.822)
Sitting	58	68.6 (11.0)	67.7 (11.0)	0.9 (5.6)	11.9, -10.1	3.8	0.865 (0.783-0.918)
Sleeping	7	69.5 (1.4)	69.4 (2.1)	—	—	—	—
Travel	22	78.1 (13.1)	79.5 (14.9)	-1.3 (9.5)	17.4, -20.1	6.5	0.764 (0.526-0.892)
Walking	20	97.4 (7.2)	90.8 (14.4)	6.6 (15.0)	36.0, -22.7	10	0.164 (-0.134-0.435)

<sup>a</sup>Number of data points analyzed for each domain.

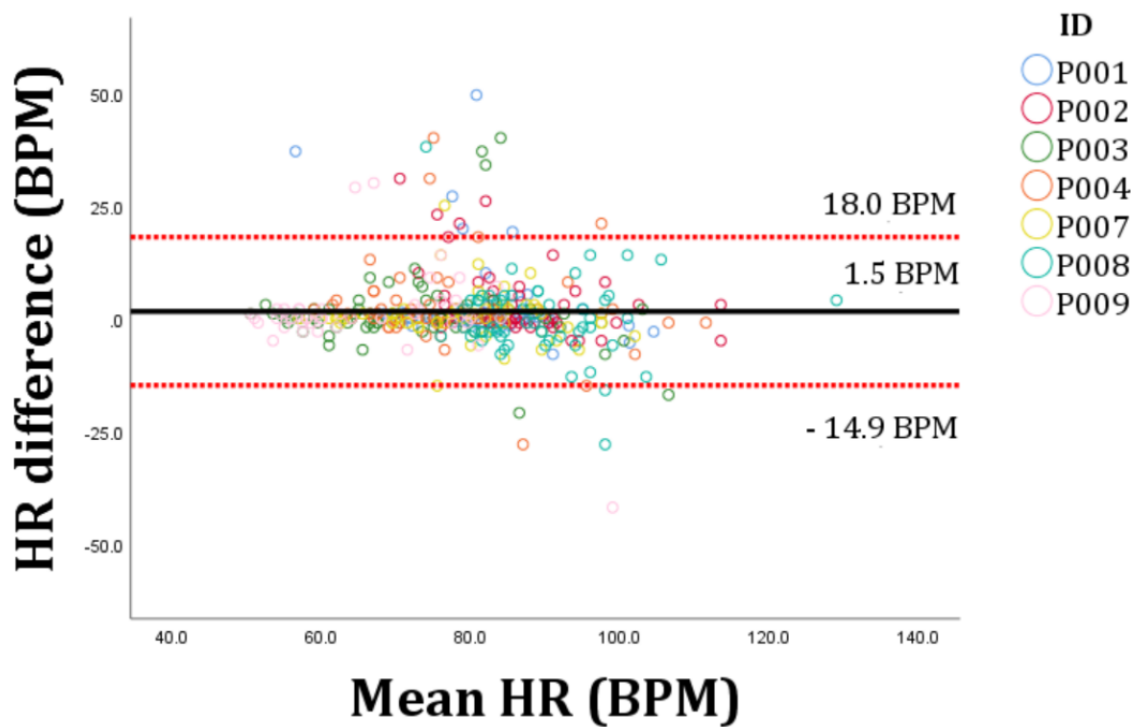
<sup>b</sup>LoA: limits of agreement.

<sup>c</sup>MAPE: mean absolute percentage error.

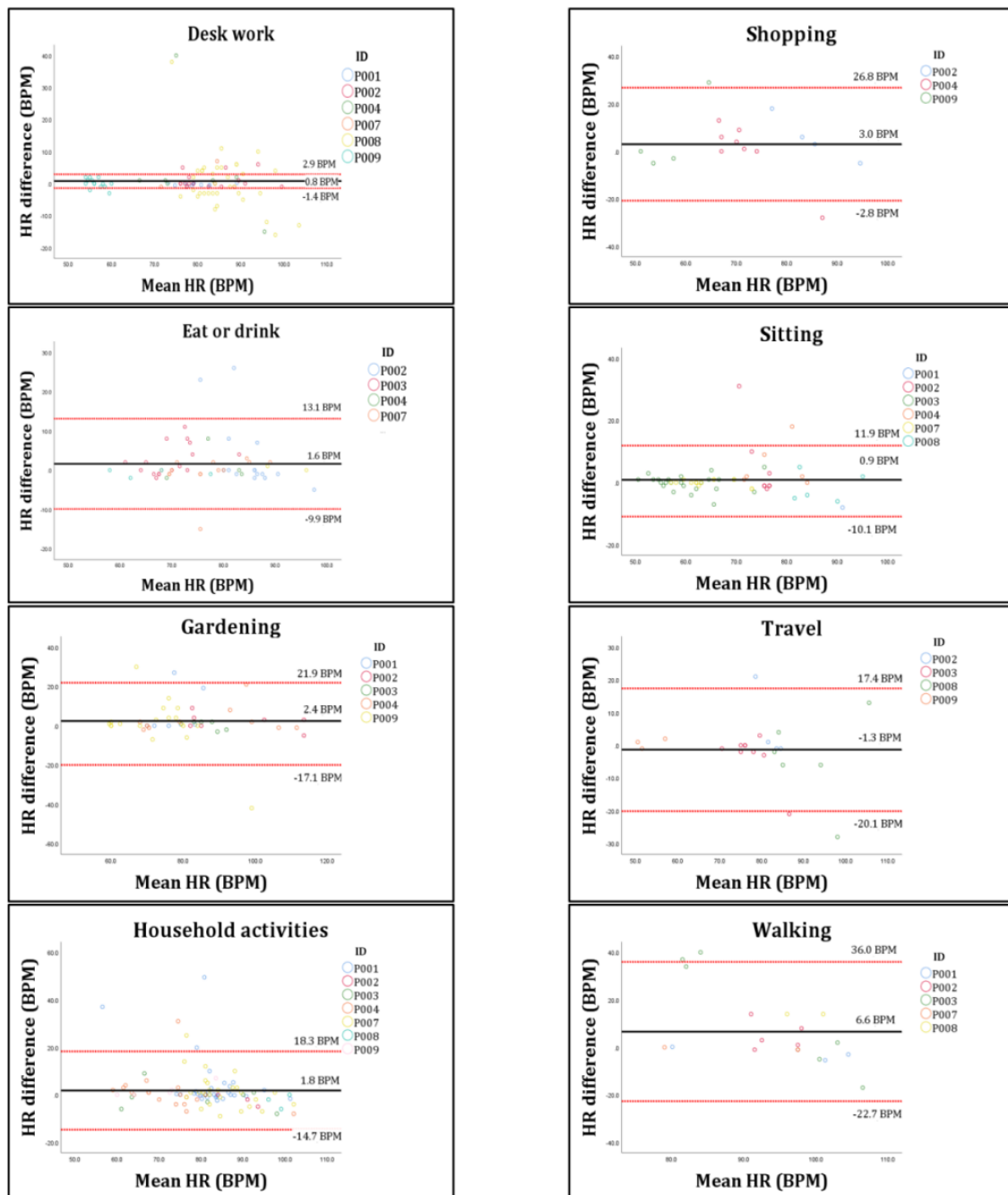
<sup>d</sup>CCC: concordance correlation coefficient.

<sup>e</sup>Not applicable.

**Figure 1.** Bland-Altman plots outlining the agreement between the Withings ScanWatch and the Polar H10 over the 12-hour study period. The full horizontal line is the bias and the dotted lines are the 95% limits of agreement. Data are displayed in BPM. BPM: beats per minute; HR: heart rate.



**Figure 2.** Bland-Altman plots outlining the agreement between the Withings ScanWatch and the Polar H10 for each activity domain. The full horizontal line is the bias and the dotted lines are the 95% limits of agreement. Data are displayed in BPM. BPM: beats per minute; HR: heart rate.



## Discussion

### Principal Findings

Wearable activity monitoring devices have attracted considerable interest over the past few years [1]. Many of these devices can now measure HR using PPG technology. Wrist-worn PPG-enabled devices have several advantages over traditional ECG-based methods of measuring HR and provide a valuable means of estimating physical activity intensity. This ecological validation study sought to assess the accuracy of HR measures produced by the Withings ScanWatch, in a sample of adult volunteers, under free-living conditions. Overall, across the 12-hour study period, the ScanWatch produced accurate measurements of HR, performing within an acceptable error

range for measuring HR (MAPE <10%) [14]. This error range is acceptable for a device intended for general consumer use; however, it would not be acceptable in settings such as health care or clinical trials where accurate measurements are paramount.

### Comparison to Prior Work

Numerous studies have examined the accuracy of HR measures produced by wrist-worn PPG-enabled devices [19-24]. A recent systematic review and meta-analysis synthesized the evidence from these studies to determine the overall validity of HR measured by such devices [13]; in this review, 15 different device brands were evaluated, with devices from Fitbit, Apple, Garmin, Mio, and TomTom being the most frequently studied. Most trials included in this review were conducted in controlled

laboratory environments with activity types such as treadmill exercise, elliptical exercise, resistance training, biking on cycle ergometer, and daily living activities being evaluated. Overall, the pooled estimates in this systematic review indicate that wrist-worn estimates of HR measurements closely resemble HR derived from the criterion measure of HR, with only two activity types—resistance training and cycling—showing sizeable differences between wrist-worn devices and the criterion measure of HR.

To the authors' knowledge, only one previous study has been conducted to assess the accuracy of HR measurements produced by devices outside laboratory and in more naturalistic environments during activities of daily living, where movements are more variable [17]. This study [17] assessed the accuracy of HR measurements produced by the Apple Watch 3 and the Fitbit Charge 3, as compared with the gold standard reference method of a 3-lead ECG, under free-living conditions, and showed that these devices were generally accurate across the 24-hour study period, with MAPE values of 5.86% and 5.96%, respectively. In our investigation, the MAPE of the ScanWatch across the 12-hour recording period was 5.1%. The ScanWatch produced accurate measurements of HR during the lower intensity activity domains of sitting, desk work, and eat or drink, with MAPE values <5% compared to the Polar H10. The ScanWatch's estimation of HR was less accurate during activities that may have involved more wrist movement. Motion artifact is commonly cited as a source of measurement error in wrist-worn PPG-enabled measurement of HR [25]. Motion artifacts have long posed a problem with the measurement of physiological signals [26], with the contamination of PPG signals usually caused by movement of the wrist or hand. In this study, participants were instructed to wear the ScanWatch on their nondominant wrist, which may have resulted in fewer erratic movements during these activities, and therefore, the impact of the motion artifact may have been limited. Future studies should examine the differences in HR measurements produced when the device is worn on dominant versus the nondominant wrist.

The results of our study revealed that the ScanWatch produced less accurate measurements of HR with increasing intensity of activity. The ScanWatch underestimated HR during the higher-intensity activity domains of gardening, household activities, walking, and shopping. The highest MAPE was observed during the walking and shopping activity domains (10% and 10.8%, respectively). These findings are similar to those of previously conducted research [27,28], which showed that increasing exercise intensity and increasing arm movement during household tasks can interfere with PPG measurement of HR. The ScanWatch overestimated HR measurements during the travel activity domain. Participants may have classified the walk to and from their mode of transport as part of the travel activity domain, thereby providing a possible explanation for the overestimation of HR measurements during the largely sedentary activity of travel. Not excluding transition periods between activities in this study, it was not possible to assess if this was the explanation for the readings. Furthermore, as with household tasks [20,21], arm movements while driving may have resulted in motion artefact that may have contaminated

readings. Further study would be required to explore this possibility.

Although the overall measurements of HR produced by the ScanWatch in the study are within an acceptable error range [17], a degree of caution should be applied when selecting this device for the measurement of HR. As the ScanWatch appears to underestimate HR during higher-intensity activities, further evaluations are recommended to examine how accurate this device is during different types and intensities of exercise. Nonetheless, for everyday activities, the ScanWatch offers a practical solution for estimating HR.

### Strengths and Limitations

To our knowledge, this is the first study examining the ecological validity of HR measurements produced by the ScanWatch, when used during daily activities, as intended for consumers in real-world contexts. However, there are some limitations to this study. First, the main limitation of this study is the difference between the ScanWatch and the Polar H10 in terms of the sampling rate employed. The Polar H10 collected data at a frequency of 1 Hz, whereas the ScanWatch sampled HR data approximately every 10 minutes. To address this, the Polar H10 data were averaged, and comparisons were performed between matched timestamps. Nonetheless, as the ScanWatch only measures HR every 10 minutes, short-term changes in HR could not be captured. This renders the device inappropriate in situations where continuous measurements of HR are required, as in health care.

Second, the study included a small sample of adults; therefore, the results cannot be generalized to any particular cohort. Future studies should include larger number of participants. In addition, the impact that demographic variables such as skin tone have on the accuracy of measurements should also be evaluated. Previous studies have shown that darker skin tones absorb more green light, which can be problematic for PPG devices using green LED light [29]. Although recent work examining the role of skin tone on wrist-worn HR devices found no significant difference in accuracy across skin tones [25], future work should continue to explore the accuracy of wrist-worn device measurements across different skin tones.

Third, due to the COVID-19 restrictions in place during the time of data collection, we had little control or oversight of how participants positioned and wore the ScanWatch and the Polar H10 devices during the study period. Participants were given written and pictorial instructions in the form of a study manual and verbal instructions via a Zoom videoconference call to guide correct device positioning, but there may have been inconsistencies in device placement between participants. Nevertheless, this reflects real-world usage of devices where the consumer independently dons a device.

Fourth, in this study, participants' self-report of activities undertaken during the 12-hour study period was used. There may have been discrepancies in how participants reported some activities and how they documented the start and end time of the activity. Previous research has demonstrated that diary recall within physical activity can result in inaccuracy and bias [30]. Future investigations could make use of body-worn cameras to

record the activities undertaken by participants during free-living activities [31]. In addition, the transition times between activities were not excluded from the data, and this may have resulted in some data points being misclassified. The free-living nature of this study resulted in some activity domains with very few data points available, and this data could only be analyzed descriptively.

Finally, the Polar H10 was used to provide the criterion measure of HR in this study. A 12-lead ECG is considered the gold standard reference measurement of HR; however, it would not have been feasible to use it in this investigation. The Polar H10 does not allow for the export of raw ECG data. The signal could therefore not be checked for noise, and this must also be acknowledged as a limitation of the study.

## Conclusions

Wrist-worn devices that incorporate PPG sensing represent an exciting means of measuring HR. This study examined the accuracy of HR measurements produced by the Withings ScanWatch during free-living activities in a sample of adult volunteers. The results revealed that the ScanWatch measured HR with acceptable accuracy during a range of day-to-day activities in a small cohort of healthy, largely sedentary, middle-aged adults. The findings indicate that this device is suitable for general consumer use. However, further investigations examining HR measurements during activities with more vigorous intensity are required before definitive conclusions on device accuracy can be made.

---

## Acknowledgments

This research is part of the Eastern Corridor Medical Engineering (ECME) project, which has been funded by European Union's INTERREG VA programme, managed by the Special EU Programmes Body (SEUPB).

---

## Data Availability

The data that support the findings of this study are available in [Multimedia Appendix 2](#).

---

## Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Study manual.

[\[DOCX File, 372 KB - formative\\_v6i9e34280\\_app1.docx\]](#)

---

### Multimedia Appendix 2

Data availability.

[\[XLSX File \(Microsoft Excel File\), 33 KB - formative\\_v6i9e34280\\_app2.xlsx\]](#)

---

## References

1. Ferreira JJ, Fernandes CI, Rammal HG, Veiga PM. Wearable technology and consumer interaction: a systematic review and research agenda. *Comput Hum Behav* 2021 May;118:106710. [doi: [10.1016/j.chb.2021.106710](https://doi.org/10.1016/j.chb.2021.106710)]
2. Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas* 2007 Mar;28(3):R1-39. [doi: [10.1088/0967-3334/28/3/R01](https://doi.org/10.1088/0967-3334/28/3/R01)] [Medline: [17322588](https://pubmed.ncbi.nlm.nih.gov/17322588/)]
3. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol* 2017 Jan 01;2(1):104-106. [doi: [10.1001/jamacardio.2016.3340](https://doi.org/10.1001/jamacardio.2016.3340)] [Medline: [27732703](https://pubmed.ncbi.nlm.nih.gov/27732703/)]
4. Weippert M, Kumar M, Kreuzfeld S, Arndt D, Rieger A, Stoll R. Comparison of three mobile devices for measuring R-R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system. *Eur J Appl Physiol* 2010 Jul;109(4):779-786. [doi: [10.1007/s00421-010-1415-9](https://doi.org/10.1007/s00421-010-1415-9)] [Medline: [20225081](https://pubmed.ncbi.nlm.nih.gov/20225081/)]
5. Andre D, Wolf DL. Recent advances in free-living physical activity monitoring: a review. *J Diabetes Sci Technol* 2007 Sep 24;1(5):760-767 [FREE Full text] [doi: [10.1177/193229680700100522](https://doi.org/10.1177/193229680700100522)] [Medline: [19885145](https://pubmed.ncbi.nlm.nih.gov/19885145/)]
6. Pereira T, Tran N, Gadhomi K, Pelter MM, Do DH, Lee RJ, et al. Photoplethysmography based atrial fibrillation detection: a review. *NPJ Digit Med* 2020;3:3 [FREE Full text] [doi: [10.1038/s41746-019-0207-9](https://doi.org/10.1038/s41746-019-0207-9)] [Medline: [31934647](https://pubmed.ncbi.nlm.nih.gov/31934647/)]
7. Selder J, Proesmans T, Breukel L, Dur O, Gielen W, van Rossum A, et al. Assessment of a standalone photoplethysmography (PPG) algorithm for detection of atrial fibrillation on wristband-derived data. *Comput Methods Programs Biomed* 2020 Dec;197:105753 [FREE Full text] [doi: [10.1016/j.cmpb.2020.105753](https://doi.org/10.1016/j.cmpb.2020.105753)] [Medline: [32998102](https://pubmed.ncbi.nlm.nih.gov/32998102/)]
8. Georgieva-Tsaneva G, Gospodinova E, Cheshmedzhiev K. Cardiodiagnostics based on photoplethysmographic signals. *Diagnostics (Basel)* 2022 Feb 05;12(2):412 [FREE Full text] [doi: [10.3390/diagnostics12020412](https://doi.org/10.3390/diagnostics12020412)] [Medline: [35204503](https://pubmed.ncbi.nlm.nih.gov/35204503/)]
9. Eerikäinen LM, Bonomi AG, Schipper F, Dekker LRC, Vullings R, de Morree HM, et al. Comparison between electrocardiogram- and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions. *Physiol Meas* 2018 Aug 08;39(8):084001. [doi: [10.1088/1361-6579/aad2c0](https://doi.org/10.1088/1361-6579/aad2c0)] [Medline: [29995641](https://pubmed.ncbi.nlm.nih.gov/29995641/)]

10. Boudreaux B, Hebert E, Hollander D, Williams B, Cormier C, Naquin M, et al. Validity of wearable activity monitors during cycling and resistance exercise. *Med Sci Sports Exerc* 2018 Mar;50(3):624-633. [doi: [10.1249/MSS.0000000000001471](https://doi.org/10.1249/MSS.0000000000001471)] [Medline: [29189666](https://pubmed.ncbi.nlm.nih.gov/29189666/)]
11. Muggeridge DJ, Hickson K, Davies AV, Giggins OM, Megson IL, Gorely T, et al. Measurement of heart rate using the Polar OH1 and Fitbit charge 3 wearable devices in healthy adults during light, moderate, vigorous, and sprint-based exercise: validation study. *JMIR Mhealth Uhealth* 2021 Mar 25;9(3):e25313. [doi: [10.2196/25313](https://doi.org/10.2196/25313)] [Medline: [33764310](https://pubmed.ncbi.nlm.nih.gov/33764310/)]
12. Khushhal A, Nichols S, Evans W, Gleadall-Siddall D, Page R, O'Doherty A, et al. Validity and reliability of the Apple watch for measuring heart rate during exercise. *Sports Med Int Open* 2017 Oct 18;1(6):E206-E211 [[FREE Full text](#)] [doi: [10.1055/s-0043-120195](https://doi.org/10.1055/s-0043-120195)] [Medline: [30539109](https://pubmed.ncbi.nlm.nih.gov/30539109/)]
13. Zhang Y, Weaver RG, Armstrong B, Burkart S, Zhang S, Beets MW. Validity of Wrist-Worn photoplethysmography devices to measure heart rate: a systematic review and meta-analysis. *J Sports Sci* 2020 Sep 19;38(17):2021-2034. [doi: [10.1080/02640414.2020.1767348](https://doi.org/10.1080/02640414.2020.1767348)] [Medline: [32552580](https://pubmed.ncbi.nlm.nih.gov/32552580/)]
14. Sartor F, Papini G, Cox LGE, Cleland J. Methodological shortcomings of wrist-worn heart rate monitors validations. *J Med Internet Res* 2018 Jul 02;20(7):e10108 [[FREE Full text](#)] [doi: [10.2196/10108](https://doi.org/10.2196/10108)] [Medline: [29967000](https://pubmed.ncbi.nlm.nih.gov/29967000/)]
15. Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. *Arch Dermatol* 1988 Jun 01;124(6):869. [doi: [10.1001/archderm.1988.01670060015008](https://doi.org/10.1001/archderm.1988.01670060015008)]
16. Etiwy M, Akhrass Z, Gillinov L, Alashi A, Wang R, Blackburn G, et al. Accuracy of wearable heart rate monitors in cardiac rehabilitation. *Cardiovasc Diagn Ther* 2019 Jun;9(3):262-271. [doi: [10.21037/cdt.2019.04.08](https://doi.org/10.21037/cdt.2019.04.08)]
17. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. *JMIR Mhealth Uhealth* 2019 Mar 11;7(3):e10828 [[FREE Full text](#)] [doi: [10.2196/10828](https://doi.org/10.2196/10828)] [Medline: [30855232](https://pubmed.ncbi.nlm.nih.gov/30855232/)]
18. 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](https://pubmed.ncbi.nlm.nih.gov/2868172/)]
19. Abt G, Bray J, Benson AC. The validity and inter-device variability of the Apple Watch™ for measuring maximal heart rate. *J Sports Sci* 2018 Jul 31;36(13):1447-1452. [doi: [10.1080/02640414.2017.1397282](https://doi.org/10.1080/02640414.2017.1397282)] [Medline: [29090987](https://pubmed.ncbi.nlm.nih.gov/29090987/)]
20. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: apple watch vs Fitbit charge HR. *J Sports Sci* 2017 Dec 06:1-8. [doi: [10.1080/02640414.2017.1412235](https://doi.org/10.1080/02640414.2017.1412235)] [Medline: [29210326](https://pubmed.ncbi.nlm.nih.gov/29210326/)]
21. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. *PLoS One* 2018;13(2):e0192691 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0192691](https://doi.org/10.1371/journal.pone.0192691)] [Medline: [29489850](https://pubmed.ncbi.nlm.nih.gov/29489850/)]
22. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. *JMIR Mhealth Uhealth* 2017 Mar 16;5(3):e34 [[FREE Full text](#)] [doi: [10.2196/mhealth.7043](https://doi.org/10.2196/mhealth.7043)] [Medline: [28302596](https://pubmed.ncbi.nlm.nih.gov/28302596/)]
23. Delgado-Gonzalo R, Parak J, Tarniceriu A, Renevey P, Bertschi M, Korhonen I. Evaluation of accuracy and reliability of PulseOn optical heart rate monitoring device. *Conf Proc IEEE Eng Med Biol Soc* 2015 Aug;2015:430-433. [doi: [10.1109/EMBC.2015.7318391](https://doi.org/10.1109/EMBC.2015.7318391)] [Medline: [26736291](https://pubmed.ncbi.nlm.nih.gov/26736291/)]
24. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photoplethysmography. *Conf Proc IEEE Eng Med Biol Soc* 2014;2014:3670-3673. [doi: [10.1109/EMBC.2014.6944419](https://doi.org/10.1109/EMBC.2014.6944419)] [Medline: [25570787](https://pubmed.ncbi.nlm.nih.gov/25570787/)]
25. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digit Med* 2020;3:18 [[FREE Full text](#)] [doi: [10.1038/s41746-020-0226-6](https://doi.org/10.1038/s41746-020-0226-6)] [Medline: [32047863](https://pubmed.ncbi.nlm.nih.gov/32047863/)]
26. Webster JG. Reducing motion artifacts and interference in biopotential recording. *IEEE Trans Biomed Eng* 1984 Dec;BME-31(12):823-826. [doi: [10.1109/tbme.1984.325244](https://doi.org/10.1109/tbme.1984.325244)]
27. Amoedo A, Martinez-Costa MDP, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: <http://los40.com>, <http://los40principales.com>, <http://cadena100.es>, <http://europafm.es> and <http://kissfm.es>. *Radio J* 2009 Feb 01;6(1):5-20 [[FREE Full text](#)] [doi: [10.1386/rajo.6.1.5\\_4](https://doi.org/10.1386/rajo.6.1.5_4)]
28. Thomson EA, Nuss K, Comstock A, Reinwald S, Blake S, Pimentel RE, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. *J Sports Sci* 2019 Jun 18;37(12):1411-1419. [doi: [10.1080/02640414.2018.1560644](https://doi.org/10.1080/02640414.2018.1560644)] [Medline: [30657025](https://pubmed.ncbi.nlm.nih.gov/30657025/)]
29. 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. *Proc Hum Factors Ergon Soc Annu Meet* 2017 Oct 20;61(1):1292-1296. [doi: [10.1177/1541931213601804](https://doi.org/10.1177/1541931213601804)]
30. Nicolson PJA, Hinman RS, Wrigley TV, Stratford PW, Bennell KL. Self-reported home exercise adherence: a validity and reliability study using concealed accelerometers. *J Orthop Sports Phys Ther* 2018 Dec;48(12):943-950. [doi: [10.2519/jospt.2018.8275](https://doi.org/10.2519/jospt.2018.8275)] [Medline: [30053792](https://pubmed.ncbi.nlm.nih.gov/30053792/)]
31. Bach K, Kongsvold A, Bårdstu H, Bardal E, Kjærnli H, Herland S, et al. A machine learning classifier for detection of physical activity types and postures during free-living. *J Meas Phys Behav* 2021 Dec 30;5(1):24-31. [doi: [10.1123/jmpb.2021-0015](https://doi.org/10.1123/jmpb.2021-0015)]

## Abbreviations

**BPM:** beats per minute  
**CCC:** concordance correlation coefficient  
**ECG:** electrocardiogram  
**HR:** heart rate  
**LoA:** limits of agreement  
**MAPE:** mean absolute percentage error  
**PPG:** photoplethysmography

*Edited by T Leung; submitted 14.10.21; peer-reviewed by M Nissen, G Jerome; comments to author 11.01.22; revised version received 26.04.22; accepted 04.08.22; published 01.09.22.*

*Please cite as:*

*Giggins OM, Doyle J, Smith S, Crabtree DR, Fraser M*

*Measurement of Heart Rate Using the Withings ScanWatch Device During Free-living Activities: Validation Study*

*JMIR Form Res 2022;6(9):e34280*

*URL: <https://formative.jmir.org/2022/9/e34280>*

*doi: [10.2196/34280](https://doi.org/10.2196/34280)*

*PMID: [36048505](https://pubmed.ncbi.nlm.nih.gov/36048505/)*

©Oonagh M Giggins, Julie Doyle, Suzanne Smith, Daniel R Crabtree, Matthew Fraser. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 01.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Digital Health Screening in People With HIV in Uganda to Increase Alcohol Use Reporting: Qualitative Study on the Development and Testing of the Self-administered Digital Screener for Health

Nneka Emenyonu<sup>1\*</sup>, MPH, DrPH; Allen Kekibiina<sup>2\*</sup>, BA, MPH; Sarah Woolf-King<sup>3\*</sup>, MPH, PhD; Catherine Kyampire<sup>2\*</sup>, BCP; Robin Fatch<sup>1\*</sup>, MPH; Carol Dawson-Rose<sup>1\*</sup>, MSc, PhD; Winnie Muyindike<sup>2\*</sup>, MBChB, MMED; Judith Hahn<sup>1\*</sup>, MA, PhD

<sup>1</sup>Division of HIV, Infectious Diseases, and Global Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, United States

<sup>2</sup>MUST Grants Office, Mbarara University of Science and Technology, Mbarara, Uganda

<sup>3</sup>Department of Psychology, Syracuse University, Syracuse, NY, United States

\*all authors contributed equally

**Corresponding Author:**

Nneka Emenyonu, MPH, DrPH

Division of HIV, Infectious Diseases, and Global Medicine

Department of Medicine

University of California San Francisco

550 16th Street

UCSF Mission Hall, 3rd Floor

San Francisco, CA, 94158

United States

Phone: 1 6505751465

Email: [nneka.emenyonu@ucsf.edu](mailto:nneka.emenyonu@ucsf.edu)

## Abstract

**Background:** Alcohol consumption is a critical driver of the HIV epidemic worldwide, particularly in sub-Saharan Africa, where unhealthy alcohol use and HIV are prevalent. Brief alcohol interventions are effective in reducing alcohol use; however, they depend on effective screening for unhealthy alcohol use, which is often underreported. Thus, there is a need to develop methods to improve reporting of unhealthy alcohol use as an essential step toward referral to brief alcohol interventions. Self-administered digital health screeners may improve reporting.

**Objective:** This study aimed to develop and test a digital, easy-to-use self-administered health screener. The health screener was designed to be implemented in a busy, underresourced HIV treatment setting and used by patients with varying levels of literacy.

**Methods:** We conducted a qualitative study at the Immune Suppression Syndrome (ISS) Clinic of Mbarara Regional Referral Hospital in Uganda to develop and test a digital self-administered health screener. The health screener included a training module and assessed behaviors regarding general health, HIV care, and mental health as well as sensitive topics such as alcohol use and sexual health. We conducted focus group discussions with clinicians and patients with HIV of the Mbarara ISS Clinic who consumed alcohol to obtain input on the need for and content, format, and feasibility of the proposed screener. We iteratively revised a tablet-based screener with a subset of these participants, piloted the revised screener, and conducted individual semistructured in-depth interviews with 20 participants who had taken part in our previous studies on alcohol and HIV, including those who had previously underreported alcohol use and with low literacy.

**Results:** A total of 45 people (n=5, 11% clinicians and n=40, 89% Mbarara ISS Clinic patients) participated in the study. Of the patient participants, 65% (26/40) were male, 43% (17/40) had low literacy, and all (40/40, 100%) had self-reported alcohol use in previous studies. Clinicians and patients cited benefits such as time savings, easing of staff burden, mitigation of patient-provider tension around sensitive issues, and information communication, but also identified areas of training required, issues of security of the device, and confidentiality concerns. Patients also stated fear of forgetting how to use the tablet, making mistakes, and losing information as barriers to uptake. In pilot tests of the prototype, patients liked the feature of a recorded voice

in the local language and found the screener easy to use, although many required additional help and training from the study staff to complete the screener.

**Conclusions:** We found a self-administered digital health screener to be appealing to patients and clinicians and usable in a busy HIV clinic setting, albeit with concerns about confidentiality and training. Such a screener may be useful in improving reporting of unhealthy alcohol use for referral to interventions.

(*JMIR Form Res* 2022;6(9):e35015) doi:[10.2196/35015](https://doi.org/10.2196/35015)

## KEYWORDS

unhealthy alcohol use; HIV; digital screening; Uganda; mobile phone

## Introduction

### Background

Alcohol consumption is a critical driver of the HIV epidemic worldwide, particularly in sub-Saharan Africa (SSA), where high-risk alcohol use and HIV are prevalent. Unhealthy alcohol use, defined as drinking more than the recommended amount of alcohol [1], is associated with increased sexual risk behavior, increased HIV transmission [2-8], and diminished treatment outcomes among people with HIV, including reduced antiretroviral (ARV) adherence [9-13] and reduced HIV viral load suppression [9,11,13-16]. Thus, reducing alcohol use among those with HIV is a public health priority.

In Uganda, more than half of the population abstains from alcohol use; however, among people who drink alcohol, most of whom are male, the yearly average consumption is 26 L of absolute alcohol, which translates to 474 L of 5.5% of alcohol by volume (typical for beer) [17]. The prevalence of heavy episodic drinking, defined as consuming at least 60 g of pure alcohol on one occasion in the previous 30 days, is 68.8% among men and 32.6% among women who consume alcohol [17]. Among people with HIV in SSA, meta-analyses have found the pooled prevalence of alcohol use disorder, defined as problem drinking that is at risk of becoming severe [18], to be 22.9% to 29.8% [19,20]. Brief interventions to reduce alcohol use have shown efficacy in reducing drinking by 15% to 30% one year after the intervention, as well as good feasibility and cost-effectiveness in primary care settings worldwide [21-28]. However, the usefulness of brief alcohol interventions depends on effective screening for unhealthy alcohol use [23,29-40], which is often underreported [41-44]. For example, we have found substantial underreporting of alcohol use by people in HIV care in Uganda to clinicians at clinic intake visits [43]. Other studies have reported high rates of underreporting of alcohol use when compared with biological measures such as phosphatidylethanol [41,44-46]. Thus, there is a need to develop methods to improve reporting of unhealthy alcohol use as an essential step toward referral to brief alcohol interventions.

### Digital Screening

The use of tablets for self-completion of clinic intake forms is increasing in many resource-rich settings; for instance, several clinical settings in the United States and Sweden have examined the use of digital screening and found that reporting of risk behaviors was comparable with more traditional screening methods [34,47-57], and the digital screening over a wide range of patient characteristics was acceptable [58-62]. Audio

computer-assisted self-interviewing (ACASI) has been associated with increased reporting of stigmatized behaviors such as forced sex, especially in Africa and rural settings [63]. ACASI has also been useful for populations with low literacy, and studies have shown high levels of comfort with its use among men and women as well as in older adults across international settings [64-66]. The rise in mobile device use worldwide makes digital screening technology ubiquitous. In addition, the simplicity of touch screen computers (tablets or smartphones) allows populations with low literacy, such as in Uganda, where 24% of the adult population (aged  $\geq 15$  years) have low literacy [67], to self-administer questionnaires that would normally require the assistance of a third party. Feasible, acceptable, and efficient methods for the assessment of sensitive behaviors in settings with low literacy are essential for comprehensive HIV care. Thus, we sought to develop and test a brief (3-5 minutes), digital, easy-to-use self-administered health screener for implementation in a busy, underresourced, low-literacy HIV treatment setting.

## Methods

### Overview

We conducted patient and clinician focus group discussions (FGDs) and in-depth interviews (IDIs) to develop and pilot-test a digital self-administered screener, called the Self-Administered Digital Screener for Health (SASH), for potential use in HIV clinic waiting rooms to increase reporting of unhealthy alcohol use in people with HIV in Uganda. We examined the acceptability of the SASH by exploring what participants thought of the health screener and their experience with the SASH, including their ability to complete it on a tablet. We also explored whether the use of the SASH would be feasible in our study setting, specifically asking questions about the practical and logistical issues of implementing the new technology within standard HIV care.

### Setting

We conducted the study at the Immune Suppression Syndrome (ISS) Clinic of the Mbarara Regional Referral Hospital (MRRH) in Uganda, an 11,000-patient clinic with 45 patients seen per day per clinician. The ISS Clinic uses electronic medical records via the Open Medical Record System [68]. Adherence counseling, which sometimes includes brief advice on alcohol use, is conducted by the HIV counselors in individual and group formats at the clinic. The MRRH is a government referral hospital in the western region of Uganda with a bed capacity of 600. It is also the teaching hospital for Mbarara University



of Science and Technology, home to the second-largest medical school in Uganda. The MRRH is located in the semirural city of Mbarara in southwestern Uganda, approximately 250 km from the capital city of Kampala. Although ISS Clinic patients come from throughout Western Uganda, our previous studies, from which the participants were sampled, limited inclusion to those who lived within 60 km or 2-hour driving distance of the ISS Clinic for ease of follow-up. Patient participants for this

study were purposefully sampled to meet study criteria from existing research databases. All clinicians attending a weekly clinic meeting were invited to participate; during the meeting, those interested were given additional study information by study staff before providing informed consent.

The study was conducted in 4 phases summarized in [Table 1](#), and as described in the following sections.

**Table 1.** Summary of procedures for the development of the Self-Administered Digital Screener for Health (SASH).

Goals and activities	Participants
<ul style="list-style-type: none"> <li>Predevelopment phase—preliminary selection of tool options and technology:               <ul style="list-style-type: none"> <li>Selection of technology most suitable for the local setting by the research team</li> <li>Review of existing literature to inform technology options and screening content</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>None</li> </ul>
<ul style="list-style-type: none"> <li>Phase 1—FGDs<sup>a</sup> to obtain initial input on the SASH:               <ul style="list-style-type: none"> <li>Obtain clinic and patient perspectives on content, format, and feasibility of use in the clinic</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Pilot: male patients (n=6, mixed literacy levels)</li> <li>Group 2: male patients (n=7, mixed literacy levels)</li> <li>Group 3: clinicians (n=5, mixed sex)</li> <li>Group 4: female patients (n=7, mixed literacy levels)</li> <li>Total of 4 FGDs with 5 clinicians and 20 patient participants</li> </ul>
<ul style="list-style-type: none"> <li>Phase 2—SASH prototype development by the study team:               <ul style="list-style-type: none"> <li>Prototype development</li> <li>Translation into Runyankole and audio recording</li> <li>Training module development</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Study team</li> </ul>
<ul style="list-style-type: none"> <li>Phase 3—iterative refinement of the SASH prototype with study participants via IDIs<sup>b</sup>:               <ul style="list-style-type: none"> <li>Obtain community, clinic, and patient perspectives on the SASH content, format, and delivery process</li> <li>Training and demonstration via study staff presenting parts of the intervention</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>First round (n=8):               <ul style="list-style-type: none"> <li>Male patients (n=5)</li> <li>Female patients (n=2)</li> <li>Clinician (n=1)</li> </ul> </li> <li>Second round (n=7):               <ul style="list-style-type: none"> <li>Male patients (n=4)</li> <li>Female patients (n=2)</li> <li>Clinician (n=1)</li> </ul> </li> <li>Total of 15 IDIs with 8 opinion leaders who emerged from the FGDs</li> </ul>
<ul style="list-style-type: none"> <li>Phase 4—pilot-testing of the SASH:               <ul style="list-style-type: none"> <li>Pilot the SASH with new patients</li> <li>Conduct qualitative IDIs to examine participants' experience using the SASH</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Group 1: patients with low or no literacy (9 male and 1 female)</li> <li>Group 2: patients who previously underreported their alcohol use (5 male and 5 female)</li> <li>Total of 20 IDIs with 20 patient participants</li> </ul>

<sup>a</sup>FGD: focus group discussion.

<sup>b</sup>IDI: in-depth interview.

### Phase 1: FGDs for SASH Development, Acceptability, and Feasibility

In phase 1, we conducted 4 FGDs to elicit input on the content, format, and feasibility of the proposed screener. All FGDs were held in a private space at the clinic. Clinicians were invited to participate after the study team gave a presentation of the study during a weekly ISS Clinic meeting. We recruited patients by sampling participants from our previous study databases [69,70] (all consented to future contact) based on previous unhealthy alcohol use underreporting, sex, and literacy levels, defined as follows. Previous underreporting was defined as not meeting self-report criteria for unhealthy drinking via the Alcohol Use Disorders Identification Test-Consumption (ie, women scoring

<3 and men scoring <4) [71] but having a phosphatidylethanol alcohol biomarker level of  $\geq 50$  ng/mL. Literacy (literate vs low or no literacy) was defined as the ability to read a prescribed sentence on a card when asked during study interviews. For this phase, we aimed to recruit 20 people who had underreported their alcohol use, including 10 people with low literacy and 10 people who were literate. We sought to include people with low literacy to represent the clinic population with low literacy. The FGD groups were balanced by literacy level and segregated by sex. Eligible participants were invited either via phone or in person during an Mbarara ISS Clinic visit.

We conducted clinician FGDs in English and patient FGDs in Runyankole, as desired by the participants. The discussions

were conducted by a research assistant (RA; AK, hereafter referred to as the RA) while another researcher (CN) recorded the sessions, took notes, and kept time. Both are Ugandan with over a decade of HIV and alcohol research experience. For this study, the RA received training in qualitative methodology from an experienced qualitative researcher on the team (SWK) and in Dedoose (SocioCultural Research Consultants, LLC) [72] by the project director (NE). Participants were asked questions about their experience discussing alcohol use at the clinic, the acceptability and feasibility of digital screening for alcohol use, suggestions for implementing digital screening in the clinic, and suggestions for the content of a screening tool. All sessions were audio recorded, transcribed, translated into English, and uploaded to Dedoose [72] for coding by the RA. Following each FGD, the RA wrote summary notes with key observations and reflections on emergent themes, tone of the discussions, and nonverbal communication such as body language and gestures. The study team analyzed the notes using a rapid approach [73].

## Phase 2: SASH Prototype Development

We based our initial set of questions on information obtained from the phase 1 FGDs (eg, which screening questions to include) and on an existing automated screening tool using interactive voice response in a study that developed and evaluated the acceptability and use of the tool in a primary care clinic waiting room [59]. The study concluded that the use of their electronic screener in the clinic setting was feasible and accepted by both clinicians and patients, with reporting rates comparable with published written questionnaires [59]. We created a tablet-based prototype of the health screener to run on the Android operating system, the most widely used operating system in Africa [74], and used CommCare (Dimagi, Inc) [75] software as it was free and user-friendly; allowed for offline use; and had capabilities for including more than one language, pictures, and audio, which were all important criteria for implementation in a limited-resource setting. On the basis of the phase 1 FGDs, we determined that we needed to include a training module. The training module, conducted on the tablet, included written and recorded instructions on how to use the device, respond to questions, select response options, repeat questions, and end the session. Two sample questions—"What year is it?" and "How old are you?"—were included as practice questions.

## Phase 3: SASH Prototype Demonstration and Modification

We modified the SASH prototype through an iterative process that comprised 2 rounds of demonstrations, participant hands-on use, and interviews, followed by a rapid analysis and modification of the prototype's content and appearance customization. For this phase, we selected 8 previous FGD participants, including people from each of the categories of respondents (clinicians, people with low or no literacy, and those underreporting unhealthy drinking), who emerged from the FGDs as opinion leaders to participate in the demonstration of the prototype. In each session, the RA demonstrated the SASH prototype, trained the participants on its use, and asked participants to use it in her presence. The RA solicited feedback on the wording and comprehensibility of the content in the local

language, Runyankole; the layout of the screen (buttons, colors, and icons); the sound or accents of the voice recordings; and the participants' preferences for a stylus or finger to touch the screen. The RA also requested feedback on the adequacy and effectiveness of the training module, which she summarized in written documents.

Following each individual demonstration and testing session, the RA wrote summary notes highlighting the user experience, content areas for discussion because of lack of clarity or other discrepancies, suggestions, questions, and comments raised by the participants. On the basis of a rapid analysis of the summary notes [73], we modified the prototype and conducted a second round of demonstrations with feedback solicitation 6 to 8 weeks later with the same participants.

## Phase 4: SASH Pilot-Testing and IDIs

In the final phase, we pilot-tested the SASH followed by IDIs with 20 new participants recruited from our previous studies and who met the criteria for unhealthy drinking, including 10 (50%) who had previously underreported their drinking and 10 (50%) with low or no literacy, as described previously.

The RA demonstrated the use of the SASH while the participants observed. Next, the participants proceeded to the training module, in English or Runyankole, on the tablet. Participants were allowed to use the training module repeatedly and with the RA's (AK) assistance as needed. The RA noted where difficulties occurred during the training module and screening questions. After completion of the SASH, participants were asked to share their experiences using the SASH in a 30- to 60-minute IDI using a semistructured interview guide. All interviews were audio recorded, transcribed, translated into English, and uploaded to Dedoose by the RA for coding and analysis. Following each IDI, the RA wrote summary notes that captured the general tone of the interview and reflected on emergent themes, nonverbal communication observed, and points for discussion with the larger study team.

## Analysis of FGDs and IDIs

Two Ugandan staff members (AK and CK) and the US-based project director (NE) initially reviewed all transcripts and written summaries from all phases of the study for completeness, language, and translation accuracy. All the documents were uploaded to Dedoose for data management and coding. The 3 individuals, all experienced in qualitative analysis, worked together to code the data concurrently using both inductive and deductive methods. A predetermined set of codes was informed by the domains of inquiry explored in the FGD and IDI guides. In addition, the analysis team open-coded transcripts and summaries to identify emerging themes. Inductive codes were defined and agreed upon by the team. Next, to ensure reliability in coding, the team triple-coded the first 4 transcripts, met as a team to ensure consistency in coding, and then completed the coding. The team worked on coded data to identify themes using thematic analysis [76,77]. We used content analysis [76,77] to interpret the individual interview data, including the systematic assignment of the predetermined codes.

## Ethics Approval

The study was approved by the University of California, San Francisco Institutional Review Board (15-6933); the Mbarara University of Science and Technology Research and Ethics Committee (02/08-15); and the Ugandan National Council for Science and Technology (HS 1977). All participants provided written informed consent before study participation in their preferred language (English or Runyankole). Transportation and refreshments were provided as incentives.

## Results

### Participants

A total of 45 people (n=5, 11% clinicians and n=40, 89% patients) participated in the study. The 5 clinicians included 1

(20%) medical officer, 1 (20%) clinical officer, 1 (20%) nurse, 1 (20%) counselor, and 1 (20%) peer educator, from whom we did not collect data on age. We attempted phone contact with 114 participants from previous National Institutes of Health-funded alcohol research studies conducted by the research team at the ISS Clinic. We reached 66 people by phone, of whom 26 (39%) declined participation and 40 (61%) enrolled in the study. Participants declined mainly because of time constraints. The 40 patient participants had a median age of 38 (IQR 32-44) years, 65% (26/40) were male, 43% (17/40) were not literate, and 70% (28/40) had previously underreported their drinking (Table 2).

**Table 2.** Patient participant demographics, Self-Administered Digital Screener for Health (SASH) study, Mbarara, Uganda (N=40).

	Participants
<b>Sex, n (%)</b>	
Female	14 (35)
Male	26 (65)
Age at SASH interview, median (IQR)	38 (32-44)
<b>Literacy, n (%)</b>	
Literate	23 (58)
No or low literacy	17 (43)
<b>Ever underreported alcohol use (AUDIT-C<sup>a</sup> negative; PEth<sup>b</sup> ≥50 ng/mL), n (%)</b>	
Yes	28 (70)
No	12 (30)
<b>Alcohol use at last study visit, n (%)</b>	
<b>Self-report</b>	
None (AUDIT-C=0)	18 (45)
Moderate (any self-report; AUDIT-C negative)	12 (30)
Hazardous (AUDIT-C positive)	10 (25)
<b>Self-report and PEth</b>	
None (no self-report; PEth <8 ng/mL)	5 (13)
Moderate (any self-report but AUDIT-C negative; 8≤PEth<50 ng/mL)	10 (25)
Hazardous (AUDIT-C positive or PEth ≥50 ng/mL)	25 (63)

<sup>a</sup>AUDIT-C: Alcohol Use Disorders Identification Test-Consumption.

<sup>b</sup>PEth: phosphatidylethanol.

## Emergent Themes in Phase 1

### Need, Usefulness, and Potential Benefits of a Health Screener

To explore the need for a health screener and the feasibility of its implementation in a busy clinic setting such as the Mbarara ISS Clinic, we conducted FGDs with clinicians and patients in which we discussed sensitive and stigmatizing health topics such as alcohol use and sexually transmitted infections. These health topics were of particular interest given the high frequency of underreporting in these areas in our previous research studies.

FGD participants noted that feelings of unease between clinicians and patients were common, which made discussions of sensitive topics challenging during routine clinic visits and, therefore, discussions of this nature were often avoided in the clinic. Difficulties stemmed from both the stigmatized nature of the topics—as a patient noted, “Me, I feel ashamed to tell health workers that I drink alcohol!”—and the underlying stresses associated with high-volume, low-resourced clinic settings. When probed about their experiences discussing sensitive health topics, patients described disrespectful treatment and worries about time:

*I fear to tell the health worker because I am worried that if I tell him this, he may insult me or respond rudely or he may decide to send me to the counselor and then I will waste all my time waiting at the counselor's place. [Male patient, aged 28 years, literate]*

Clinicians unanimously described their experience with asking patients about substance use as frustrating and difficult as they felt that the patients were not truthful:

*Many of them actually tell us lies if we are asking them if they drink alcohol! [Male clinician, medical officer]*

Encounters with patients who consume alcohol were described in the same punitive tone by all clinicians, as were their perceptions of alcohol-induced behavior, which included nonadherence to medications and risky sexual behavior among their patients who drank alcohol. Patients often feared being negatively judged if they disclosed behavior against which they had been advised by clinicians. Averting blame or punitive action was the main reason cited for not engaging in dialogue with clinicians, with whom they wanted to maintain a sense of dignity and respect:

*...sometimes you go there and you really want to talk about alcohol with them, and you know, they tell us that if you are taking ARVs you should stop taking alcohol, but when you come to the clinic, knowing that you took alcohol the previous day, and you are sent to the counselor, you will not tell the counselors that you took alcohol because of the fear that they will insult you or you fear them changing their attitude towards you, so you keep quiet about alcohol or at the end of it all they will tell you to get out of their sight because they will say that for you if you decided to take alcohol instead of ARVs, there is no reason as to why they should listen to you... [Male patient, aged 37 years, low literacy]*

In addition, many patients felt restricted by limited time with clinicians during visits and felt that the clinicians were overworked and fatigued. Time constraints further exacerbated communication gaps and led to the prioritization of primary HIV care concerns such as medication refills. Additional delays by talking with counselors meant to some patients that “chances are high that you will go back home without medication,” for instance, in the event that the pharmacy stock ran out or the pharmacy closed; therefore, these discussions were avoided completely to ensure medication refills.

### **Privacy Concerns, Communication, and Machine Errors That Affect the Use of a Health Screener**

Clinicians and patients both felt that a digital screener could aid in difficult conversations by enabling the disclosure of health issues that would otherwise be avoided in face-to-face clinic encounters. Patients expressed interest in receiving feedback on their alcohol use from the tablet itself. Patients felt that the use of a nonjudgmental device would help them articulate their issues more clearly to enable a better understanding of their problems:

*I also see that it will make me happy, because the tablet will send out information that I have said, it cannot give information that I have not said. And because we never get an opportunity to explain our problems to the doctors, by the time they read my concerns from the tablet, they will know that it was really a problem. [Female patient, aged 33 years, low literacy]*

Younger participants (aged <40 years) with higher literacy preferred the tablet to a health care worker, particularly for discussing sensitive topics:

*If I am talking to this computer, I will be comfortable, without looking at someone's face and how they change their face whenever they are not happy with anything you did; I will not be looking at someone to be intimidated by their facial expression. Because the computer tablet asked me very well that: have you used any other medication or herbs or if you have any other sickness and then you reply yes or no. Isn't that easy telling this to a doctor or nurse who will not listen to you but who will just be rude to you. [Male patient, aged 30 years, low literacy]*

Clinicians and patients both felt that a digital screener could be a useful clinic tool if they had access to reports from it before seeing patients. Facilitators to implementing such a tool in the ISS Clinic included expected decreases in patient wait times and reduced clinician workload, which would ultimately improve patient care. Clinicians felt that the use of a digital screener would also allow for more anonymity and comfort that would lead to more accurate and comprehensive disclosure:

*I think they can share, no one is there, and they are not going to be penalized for the information they have given. So, I think they will be comfortable to give, to open up, they will open up about their life. [Female clinician, clinical officer]*

Despite the comfort levels expressed in the use of a digital device, there were concerns about loss of privacy and irreparable machine errors that could affect patient care expressed by both clinicians and patients, as well as the need to physically secure the tablet in the clinic to prevent loss:

*In the event that people who keep this device lose it, a bad person may pick and then share the information with wrong people who may spread information about you. [Male patient, aged 56 years, literate]*

We asked the participants what they thought should be done with their answers to the screening questions. Participants' reactions ranged from preferring that a printed paper be included in their medical records to preferring that SMS text messages or emails be sent electronically to clinicians. Concerns about electronic transmission centered on privacy and protection of personal data or unreliable network issues that would either delay or fail delivery to clinicians. Some patients thought that the device could increase communication with their providers but worried that technical problems could send the wrong information to clinicians.

### ***Concerns About Time Constraints, Logistics, and Training Required to Implement a Health Screener***

In response to our inquiries on the feasibility of implementing a digital screener in the clinic setting, participants expressed more practical concerns regarding time, cost, space, and effort required as prohibitive. Some participants felt that a digital screener could save time during clinic visits and ease clinician fatigue, as illustrated by an older female patient:

*I think that this computer tablet is fast. You see, when you go to see a health worker then he has to ask you and this takes some time, while others are waiting to see the same doctor for the same services. But if you are using this tablet, it becomes easier and quicker compared to talking to a health worker. [Female patient, aged 69 years, literate]*

The aforementioned enthusiasm was tempered by clinicians' apprehension about the additional time and effort needed for training and additional logistical challenges involved in implementing new technology on top of their heavy workload. Specifically, clinicians worried about added time, staffing, and clinic flow disruptions that would result from added training in a clinic setting that was already understaffed with limited resources. This concern was also voiced by patients who worried about their ability to understand and operate a novel digital device.

#### **Phase 2**

The first version of the SASH prototype was developed by the study team (phase 2). We considered 5 topics (pain, smoking,

alcohol use, physical activity, and depressed mood) used in a previous waiting room screener [59] plus ARV adherence. The chosen topics for inclusion in the SASH were agreed upon by the study team and were based on what we considered to be the most relevant in our study setting from our previous experience in the Mbarara ISS Clinic. It was also important for the study team to screen for behaviors that would allow for feasible interventions in the future in this setting. We included questions about ARV adherence in the SASH as it was frequently discussed during the FGDs and we felt that including general health questions along with questions on sensitive behaviors would help normalize the SASH as a general clinic tool, making it more acceptable to patients in this setting where sensitive behaviors such as alcohol use are often stigmatized. We used validated questions on adherence [78], depression [79], and alcohol use [71].

The final content of the SASH included a training module with 2 questions plus 13 questions that covered sex, general health status, adherence to ARVs and other medicines and herbs, alcohol use, sexual health, and depression, as well as a final question asking participants if they wanted to talk to a clinician about their health issues (Table 3). The responses to this last question, as with others in the SASH, were not followed by clinic referrals as the scope of the study was limited to testing the SASH. The SASH included color-coded multiple-choice questions and response options (Figure 1), with an additional option for participants to play the corresponding audio clip by pressing the audio symbol shown. The entire SASH was available in English or Runyankole, the local language.

**Table 3.** Self-Administered Digital Screener for Health (SASH) training module and screener topics, questions, and response options.

SASH topics and sections	SASH instructions and questions	SASH response options
<b>SASH training module</b>		
Introduction	Hello! We are testing out this small computer for possible use in the clinic. First, we want to show you how to use this small computer. If you want, you can listen to someone read each question out loud. To hear a question, click on the picture of the speaker on the right. Whenever you see the picture of the speaker, you can touch it to hear the question read out loud. You do not have to touch it very hard, but do give it a little tap. Tap it again if you want to stop the voice. Please tap on the speaker picture now to hear a message.	N/A <sup>a</sup>
Example 1a	This small computer will ask you questions. The answers to these questions will sometimes have a picture or a color connected to them. The small computer will not take your picture. To answer the question, you will tap on the picture or color of the choice that you feel answers the question the best. Please tap the arrow to try an example question.	N/A
Example 1b	What year is it?	<ul style="list-style-type: none"> <li>● 1997 (red)</li> <li>● 2007 (black)</li> <li>● 2017 (yellow)</li> </ul>
Example 2a	Great! To type numbers into the small computer, you will tap the numbers that you will see at the bottom of the screen. When you are done tapping the numbers, you will need to tap the arrow to move to the next page. Please tap the arrow to try an example question.	N/A
Example 2b	How old are you?	N/A
Example 3a	Good! Please tap on the green bar above to finish the training.	N/A
<b>SASH screener</b>		
Introduction	Hello! Please answer each question as best you can.	N/A
Sex	Are you a man or a woman?	<ul style="list-style-type: none"> <li>● Man (image)</li> <li>● Woman (image)</li> </ul>
Health	In general, would you say your health is today:	<ul style="list-style-type: none"> <li>● Excellent (red)</li> <li>● Very good (yellow)</li> <li>● Good (green)</li> <li>● Fair (black)</li> <li>● Poor (purple)</li> </ul>
HIV medication adherence	Have you had any trouble taking your ARVs <sup>b</sup> lately?	<ul style="list-style-type: none"> <li>● Yes (green check)</li> <li>● No (red X)</li> </ul>
HIV medication adherence	How have you been at taking your ARVs in the last 4 weeks?	<ul style="list-style-type: none"> <li>● Excellent (red)</li> <li>● Very good (yellow)</li> <li>● Good (green)</li> <li>● Fair (black)</li> <li>● Poor (blue)</li> <li>● Very poor (purple)</li> </ul>
Other medications	Are you currently taking any herbs or medicines other than those you are given here at the clinic?	<ul style="list-style-type: none"> <li>● Yes (green check)</li> <li>● No (red X)</li> </ul>
Alcohol use	Have you taken any alcohol in the past 3 months?	<ul style="list-style-type: none"> <li>● Yes (green check)</li> <li>● No (red X)</li> </ul>
Alcohol use	On how many days have you had at least one drink of alcohol in the last 4 weeks?	<ul style="list-style-type: none"> <li>● 3 or more days per week (green)</li> <li>● 1-2 days per week (yellow)</li> <li>● 2-3 times in the past 4 weeks (purple)</li> <li>● 1 time in the past 4 weeks (red)</li> <li>● Never in the past 4 weeks (blue)</li> </ul>

SASH topics and sections	SASH instructions and questions	SASH response options
Alcohol use	How often did you have 6 or more drinks of alcohol on one occasion in the past 3 months?	<ul style="list-style-type: none"> <li>• Never (red)</li> <li>• Monthly (blue)</li> <li>• Weekly (black)</li> <li>• Daily or almost daily (yellow)</li> </ul>
PHQ <sup>c</sup>	Over the past 2 weeks, how often have you been bothered by having little or no interest in doing things?	<ul style="list-style-type: none"> <li>• Not at all (red)</li> <li>• Several days (black)</li> <li>• More than half the days (purple)</li> <li>• Nearly every day (yellow)</li> </ul>
PHQ	Over the past 2 weeks, how often have you been bothered by feeling down, depressed, or hopeless?	<ul style="list-style-type: none"> <li>• Not at all (red)</li> <li>• Several days (black)</li> <li>• More than half the days (purple)</li> <li>• Nearly every day (yellow)</li> </ul>
STIs <sup>d</sup>	Have you had any symptoms of a sexually transmitted infection recently?	<ul style="list-style-type: none"> <li>• Yes (green check)</li> <li>• No (red X)</li> <li>• Don't know (black question mark)</li> </ul>
Other discussion with health worker	Is there anything else you might want to talk to a health worker about?	<ul style="list-style-type: none"> <li>• Yes (green check)</li> <li>• Check No (red X)</li> </ul>
Other discussion with health worker	Thank you for answering all of these questions! Do you want to talk to a health worker about these things today?	<ul style="list-style-type: none"> <li>• Yes (green check)</li> <li>• Check No (red X)</li> </ul>
Other discussion with health worker	Would you like us to text a health worker or print out your information on a piece of paper for you to give them?	<ul style="list-style-type: none"> <li>• Print (yellow)</li> <li>• Text (red)</li> <li>• Both (purple)</li> <li>• No preference (black)</li> </ul>
Final thank you	Thank you! You have answered all of the questions. Please tap on the green bar above to finish the screener.	N/A

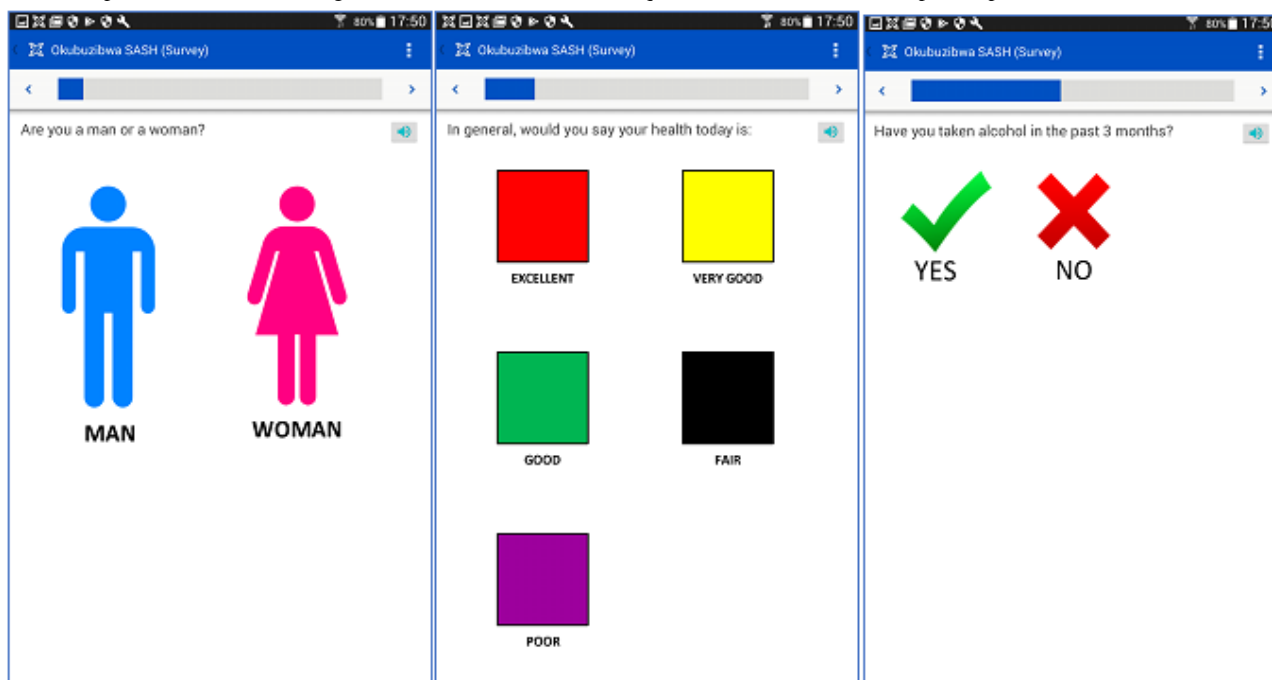
<sup>a</sup>N/A: not applicable; response not required.

<sup>b</sup>ARV: antiretroviral.

<sup>c</sup>PHQ: Patient Health Questionnaire Depression Scale.

<sup>d</sup>STI: sexually transmitted infection.

**Figure 1.** Sample Self-Administered Digital Screener for Health (SASH) questions with color-coded response options.



### Emergent Themes in Phase 3: Additional Training and Modifications to Screener Questions

The SASH was further refined through an iterative process with the patients (phase 3). Following the demonstrations of the SASH, patients provided input on the content, order of questions, and mode of delivery. We determined that we needed to allow for multiple training sessions after all participants requested additional training time and repeated the training module several times. Thus, we separated the training module from the actual screening questions to allow participants to repeat the training module until they felt comfortable with and confident about the technology. The inclusion of multiple-choice as opposed to open-ended response options, particularly with questions on the “number of days drinking,” was also guided by feedback from patients who felt that it was difficult to recall these measures with accuracy. Both clinicians and patients in the FGDs felt that it was best to place perceived difficult questions, such as those about depression, at the end. Some suggestions were outside the scope of a screening tool (eg, a female patient suggested that the digital screener serve as a medication adherence intervention).

### Emergent Themes in Phase 4

#### *Time and Comprehension Needed to Complete the SASH*

In this phase, it took 20 participants approximately 20 minutes to complete the training module and 30 minutes to complete the screener. The RA noted that a few participants took >1 hour to complete the screener as they requested additional training sessions and demonstrations with the RA. When given the option to use audio or read the screener, the RA noted that more than half of the participants (14/20, 70%) chose the audio option, including 15% (3/20) who were literate.

#### *General Satisfaction, Feelings of Empowerment, and Self-reflection With the SASH*

Participants testing the prototype in phase 4 reacted positively to the SASH. They liked it, found it easy to use, and found the local language (Runyankole) option appealing, as expressed by one of them:

*What was so easy is the fact that I could use it in my language, Runyankole.* [Male patient, aged 41 years, literate]

Participants appreciated the short length of the screener, simplicity of the screening questions and response options, and clarity of the instructions. A female participant with low literacy explained that “it tells and guides you on what to do and that’s what I liked,” and others found this particularly helpful in building confidence to proceed from the training module to the screening questions with ease:

*I have liked it because it has short questions and they explain to the point. You read the question and you quickly understand it and you find a way of answering it. And each question has its own answer.* [Male patient, aged 56 years, literate]

The availability of multiple choices for responses was empowering to some:

*The instructions made me happy because they gave me an option if I did not like something, then I would choose something else altogether. It shows different responses for example if one is weak or strong, sick or not sick, so one chooses accordingly. It shows what you are, it does not disagree with you but lets you choose.* [Female patient, aged 39 years, literate]

The audiovisuals and colorful images simplified the experience and eased comfort levels, especially for participants with low literacy and limited experience with technology:



*They were good, the fact that there are pictures, if one knew how to read, all would be fine. But for some of us that don't know how to read, you can't comprehend, you just see the pictures and try to follow...* [Female patient, aged 43 years, low literacy]

A male participant with low literacy likened the SASH to his smartphone: "I did not fear because I use a smart phone every day." Another female participant with low literacy commented on how accomplished she felt in using such novel technology:

*I felt good, when I saw that it is a touch screen, I wondered, is this the technology that I have been hearing about. I then said to myself, let me see it. When I get home, I will narrate to people I used technology that I have never used before.* [Female patient, aged 43 years, low literacy]

In some cases, answering questions prompted patients to self-reflect on their health and behavior. In addition to self-reflection, some participants felt that the SASH held them accountable and compelled honesty in their communication about their health with their providers:

*Devices that keep secrets like this one...if an expert opens it and goes through it, he will not fail to get something. So better you speak the truth so that in future, it does not reveal your lies or report you.* [Male patient, aged 49 years, literate]

### **Training Concerns and Suggestions for Improvement of the SASH**

Many patients were concerned about making mistakes in their responses or forgetting instructions while completing the SASH. We noticed this when demonstrating the SASH to participants and also during the training module. However, all participants, including those with low literacy, were open and eager to use the technology with adequate instruction and guidance. Participants felt that multiple training sessions and ongoing instruction would facilitate the implementation of the SASH:

*...but it's important to come back and repeat the training. Even those that study cannot study just once and finish.* [Male patient, aged 44 years, low literacy]

Although it took participants much longer than expected to complete the SASH (approximately 1 hour vs 3-5 minutes), particularly for those who took over an hour, it seemed that the time spent on the SASH was less of a concern to participants than their concerns about being accurate and thorough. Some participants expressed worries about forgetting how to use the tablet and making mistakes and feared the loss of information and damage to the device. Only a small number of participants voiced the need for improvements. In all, 10% (2/20) of female participants with low literacy suggested adding a video to enable a more personable experience with the nonhuman device.

Participants had varied understandings of the functions of the SASH. An ancillary finding during analysis was that our translation of "digital screener" was not precise because of the limited vocabulary in the local language for a specific description. Therefore, "digital screener" was translated to a more general term such as "technology" in Runyankole. Some

participants clearly understood the SASH as a screener, whereas others had broader expectations of its function and ability to provide health care. The latter group sometimes viewed the SASH as an educational tool, a digital suggestion box for lodging complaints, a tool to intervene on alcohol use and other behavioral issues, or even as a complete replacement for health care providers:

*It taught me to completely stop alcohol and then I will have good health. Now if you go by the advice given by this technology, it educates you and if you go by the rules, you become healthy and you live a longer life.* [Male patient, aged 52 years, literate]

After the patients used the SASH, we asked them about their preference for alcohol use screening with the SASH versus a clinician. Participants reported a range of responses, from the preference for SASH to clinician preference. Those who preferred the SASH expected that it would spare them the discomfort of dealing with difficult encounters with clinicians as well as maximize their time at the clinic:

*It is better than the provider because the provider will get out of her mood and then he shouts at you but as for the technology, it will never get out of its mood. You input what you want and then the provider will read it. The technology doesn't act in a mean way.* [Male patient, aged 41 years, literate]

The participants who preferred in-person interactions with clinicians felt that the SASH lacked flexibility in responding to issues not already programmed or in responding at all. This was coupled with concerns about grasping the technology:

*You see I do not see the person talking in the tablet, yet for the clinician we are face to face. For example, If I had a wound, the tablet would not be able to see it or prescribe medication for me, yet the clinician can do it. The tablet cannot bandage my hand.* [Female patient, aged 42 years, low literacy]

For some, the preference for a digital tool versus direct contact with a clinician was clearly not an issue as they expressed the expectation of similar outcomes for both:

*There is no difference between the two. When the clinician asks you, you answer him and when the device asks you, you have to give an answer. So, no difference.* [Female patient, aged 56 years, low literacy]

## **Discussion**

### **Principal Findings**

We developed and pilot-tested a touch screen, digital health screener with the potential to increase reporting of unhealthy alcohol use in people with HIV in Uganda, a low-resource setting with varying levels of literacy and reasons for underreporting alcohol use. The resulting SASH is a health screener with 15 screening questions illustrated with colorful images and a voice option that reads the questions and response options to the participants. We found that the SASH was acceptable to clinicians and patients who consumed unhealthy

levels of alcohol, including those with a history of underreporting their alcohol use, and usable by patients with a range of literacy levels. Although our scope did not include feasibility studies, we explored patient and clinician perceptions of the feasibility of implementing a digital screener in their clinic setting. Their responses focused on practical concerns such as where and when to use the tablet within the clinic; charging the tablet; cost; time taken to complete the screener, including training; and the potential additional workload for clinic staff who would have to train patients to use the tablet and screener. Clinicians and patients in our study shared a desire for the SASH to be used to improve patient care through data sharing with clinicians to mitigate communication barriers between patients and clinicians and save clinicians' time.

Training, literacy, and privacy of information were the primary concerns regarding the use of a digital screener in this low-resource and low-literacy setting. We found that including a training module preceding the screening questions was crucial. The fear of "failing" the screening or mishandling the device made participants particularly anxious, and all (20/20, 100%) requested retraining with the RA during testing because of this. It was unclear whether these concerns about breaking the device while handling it were rooted in patients' general concerns about the punitive consequences of making mistakes or mere fear of using new technology. Nonetheless, a notable finding in all phases of this study was that additional training beyond the basic instructions and demonstration of the prototype, including repeated practice sessions, was critical for many clinicians and patients. Training was emphasized as essential regardless of the participants' previous experience with smartphones, electronic devices, or other technology. The need for additional training, as described previously, was similarly noted in another study of new technologies in a clinic setting in SSA [80], as well as in previously mentioned studies that tested ACASI in settings with older adults and patients with low literacy [64-66].

### Comparison With Prior Work

Our results are consistent with previous studies that demonstrated the feasibility and acceptability of digital and web-based previsit screening for substance use [59,81-87]. Although most of these studies were conducted in high-income countries, the results of those conducted in SSA are promising as well [88-92]. A meta-analysis has shown that digital screening followed by brief alcohol interventions is effective in reducing weekly drinking [93]; however, few studies have focused on whether digital screening increases entry into interventions. By contrast, this study aimed to develop a screening tool as a critical precursor for intervention in a setting where alcohol use is both prevalent and stigmatized [43].

We found evidence of tensions between health care providers and patients that affected their ability to discuss sensitive health topics, including sexual health and substance use, during routine clinic visits. This is consistent with a qualitative study with patients and providers in a rural primary care clinic in the United States that found patient-provider relationships to be critical in the feasibility of substance use screening and that patients' preference was for self-administered, tablet-based screeners [94]. Clinicians in our study expected patients to underreport

their drinking but also expressed frustration when patients were not forthcoming about their drinking on consultation. Similarly, patients frequently described their interactions with providers as hostile, with punitive consequences for behaviors considered to be unacceptable, including unhealthy alcohol use. Therefore, avoidance of punitive consequences was a reason for underreporting in addition to maintaining dignity in the presence of health care providers. The preservation of social status for people with HIV, already burdened by the stigma of HIV, was noted as a reason for socially desirable reporting in this study. Both clinicians and patients felt that clinicians did not have enough time for patients during clinic visits and, as such, deeper explorations of mental health and behavioral issues in the context of current HIV care were limited. These findings suggest that digital screening methods may help mitigate the barriers imposed by face-to-face questioning on sensitive topics. More private and anonymous screening methods may prove particularly useful in settings where sensitive behaviors are stigmatized and pressures within patient-provider relationships prohibit accurate self-reporting.

Our success in developing and pilot-testing the SASH with 20 people with HIV who drank alcohol and had varied levels of literacy in the Mbarara ISS Clinic could potentially be replicated and scaled up in similar settings. Although this screener primarily focused on alcohol use, waiting room screening tools have the potential to efficiently and sensitively screen for key health issues such as mental health and medication adherence and ultimately lead to improved referral and treatment for several health and psychosocial issues.

### Limitations

This study had some limitations. The time required to train patients to use a new digital device may have diminished the focus on its actual use as a screener and may have negatively influenced perceptions. We intended for the SASH to be brief (3-5 minutes), but it took participants a relatively long time (approximately 1 hour or more) to complete the 15-question screener with a training module. Furthermore, our translation of "digital screener" more generally as "technology" in Runyankole may have affected patients' understanding of the larger context of the SASH as a screener and limited their input on content development as well as their discussion of its specific use as a screener in the clinic. Therefore, feedback from some patients focused on the use of the SASH as a learning tool, medical resource, or digital suggestion box to which complaints about their care could be lodged rather than on its use solely as a screener. It was beyond the scope of this study to provide the SASH results to clinicians and determine their impact on patient care or clinical practice. Finally, given our limited scope, we could not fully explore the feasibility of implementing the SASH in this clinic setting.

### Conclusions

Our findings suggest that a digital health screener has the potential to improve reporting of unhealthy alcohol use in clinical care for referral to alcohol interventions in HIV clinics in low-resource settings. Further studies are needed to determine the efficacy of the SASH in improving self-reporting and further develop means of implementation.

## Acknowledgments

The authors would like to thank the Mbarara University of Science and Technology; the Mbarara Regional Referral Hospital; the staff of the Mbarara Regional Referral Hospital Immune Suppression Syndrome Clinic; and our research and administrative teams at University of California, San Francisco and Mbarara University of Science and Technology. They give special thanks to the study participants for their time and contribution to this work. This research was supported by the National Institutes of Health through the National Institute on Alcohol Abuse and Alcoholism (K24 AA022586; principal investigator: JH) and the National Institute of Allergy and Infectious Diseases (4P30AI027763-25; principal investigator: Monica Gandhi).

## Authors' Contributions

JH conceived and designed the study. SWK, NE, RF, AK, and WM contributed to the study design, data analysis and interpretation, and writing of the manuscript. CK contributed to data analysis and interpretation and writing of the manuscript. CDR critically reviewed the manuscript. All authors read and approved the final version of the manuscript and take responsibility for data integrity and accuracy of the analysis.

## Conflicts of Interest

JH received consulting fees from Pear Therapeutics in 2022.

## References

1. Jin J. Screening and counseling to reduce unhealthy alcohol use. *JAMA* 2018 Nov 13;320(18):1948. [doi: [10.1001/jama.2018.16735](https://doi.org/10.1001/jama.2018.16735)]
2. Fisher J, Bang H, Kapiga S. The association between HIV infection and alcohol use: a systematic review and meta-analysis of African studies. *Sex Transm Dis* 2007;34(11):856-863. [doi: [10.1097/olq.0b013e318067b4fd](https://doi.org/10.1097/olq.0b013e318067b4fd)]
3. Kalichman SC, Simbayi LC, Kaufman M, Cain D, Jooste S. Alcohol use and sexual risks for HIV/AIDS in Sub-Saharan Africa: systematic review of empirical findings. *Prev Sci* 2007 Jan 31;8(2):141-151. [doi: [10.1007/s11121-006-0061-2](https://doi.org/10.1007/s11121-006-0061-2)]
4. Chersich MF, Bosire W, King'ola N, Temmerman M, Luchters S. Effects of hazardous and harmful alcohol use on HIV incidence and sexual behaviour: a cohort study of Kenyan female sex workers. *Global Health* 2014;10(1):22. [doi: [10.1186/1744-8603-10-22](https://doi.org/10.1186/1744-8603-10-22)]
5. Musinguzi G, Bwayo D, Kiwanuka N, Coutinho S, Mukose A, Kabanda J, et al. Sexual behavior among persons living with HIV in Uganda: implications for policy and practice. *PLoS ONE* 2014 Jan 23;9(1):e85646. [doi: [10.1371/journal.pone.0085646](https://doi.org/10.1371/journal.pone.0085646)]
6. Mehra D, Östergren P, Ekman B, Agardh A. Inconsistent condom use among Ugandan university students from a gender perspective: a cross-sectional study. *Global Health Action* 2014 Apr 10;7(1):22942. [doi: [10.3402/gha.v7.22942](https://doi.org/10.3402/gha.v7.22942)]
7. Kerridge BT, Tran P, Hasin DS. Intoxication at last sexual intercourse and unprotected sex among HIV-positive and HIV-negative individuals in Uganda: an event-level analysis. *AIDS Behav* 2014 Jul 30;19(3):412-421. [doi: [10.1007/s10461-014-0854-4](https://doi.org/10.1007/s10461-014-0854-4)]
8. Rehm J, Probst C, Shield KD, Shuper PA. Does alcohol use have a causal effect on HIV incidence and disease progression? A review of the literature and a modeling strategy for quantifying the effect. *Popul Health Metrics* 2017 Feb 10;15(1):4. [doi: [10.1186/s12963-017-0121-9](https://doi.org/10.1186/s12963-017-0121-9)]
9. Parsons JT, Rosof E, Mustanski B. The temporal relationship between alcohol consumption and HIV-medication adherence: a multilevel model of direct and moderating effects. *Health Psychol* 2008 Sep;27(5):628-637 [FREE Full text] [doi: [10.1037/a0012664](https://doi.org/10.1037/a0012664)] [Medline: [18823189](https://pubmed.ncbi.nlm.nih.gov/18823189/)]
10. Braithwaite RS, McGinnis KA, Conigliaro J, Maisto SA, Crystal S, Day N, et al. A temporal and dose-response association between alcohol consumption and medication adherence among veterans in care. *Alcohol Clin Exp Res* 2005 Jul;29(7):1190-1197. [doi: [10.1097/01.alc.0000171937.87731.28](https://doi.org/10.1097/01.alc.0000171937.87731.28)] [Medline: [16046874](https://pubmed.ncbi.nlm.nih.gov/16046874/)]
11. Weiser S, Palar K, Frongillo E, Tsai A, Kumbakumba E, Depee S, et al. Longitudinal assessment of associations between food insecurity, antiretroviral adherence and HIV treatment outcomes in rural Uganda. *AIDS* 2014 Jan 02;28(1):115-120 [FREE Full text] [doi: [10.1097/01.aids.0000433238.93986.35](https://doi.org/10.1097/01.aids.0000433238.93986.35)] [Medline: [23939234](https://pubmed.ncbi.nlm.nih.gov/23939234/)]
12. Velloza J, Kemp CG, Aunon FM, Ramaiya MK, Creegan E, Simoni JM. Alcohol use and antiretroviral therapy non-adherence among adults living with HIV/AIDS in sub-Saharan Africa: a systematic review and meta-analysis. *AIDS Behav* 2020 Jun 31;24(6):1727-1742 [FREE Full text] [doi: [10.1007/s10461-019-02716-0](https://doi.org/10.1007/s10461-019-02716-0)] [Medline: [31673913](https://pubmed.ncbi.nlm.nih.gov/31673913/)]
13. Hendershot C, Stoner S, Pantalone D, Simoni J. Alcohol use and antiretroviral adherence: review and meta-analysis. *J Acquir Immune Defic Syndr* 2009;52(2):180-202. [doi: [10.1097/qai.0b013e3181b18b6e](https://doi.org/10.1097/qai.0b013e3181b18b6e)]
14. Nakimuli-Mpungu E, Bass JK, Alexandre P, Mills EJ, Musisi S, Ram M, et al. Depression, alcohol use and adherence to antiretroviral therapy in sub-Saharan Africa: a systematic review. *AIDS Behav* 2012 Nov 25;16(8):2101-2118. [doi: [10.1007/s10461-011-0087-8](https://doi.org/10.1007/s10461-011-0087-8)] [Medline: [22116638](https://pubmed.ncbi.nlm.nih.gov/22116638/)]

15. Haberer JE, Kahane J, Kigozi I, Emenyonu N, Hunt P, Martin J, et al. Real-time adherence monitoring for HIV antiretroviral therapy. *AIDS Behav* 2010 Dec 31;14(6):1340-1346 [FREE Full text] [doi: [10.1007/s10461-010-9799-4](https://doi.org/10.1007/s10461-010-9799-4)] [Medline: [20809380](https://pubmed.ncbi.nlm.nih.gov/20809380/)]
16. Williams EC, Hahn JA, Saitz R, Bryant K, Lira MC, Samet JH. Alcohol use and human immunodeficiency virus (HIV) infection: current knowledge, implications, and future directions. *Alcohol Clin Exp Res* 2016 Oct 22;40(10):2056-2072 [FREE Full text] [doi: [10.1111/acer.13204](https://doi.org/10.1111/acer.13204)] [Medline: [27696523](https://pubmed.ncbi.nlm.nih.gov/27696523/)]
17. Global Status Report on Alcohol and Health 2018. Geneva: World Health Organization; 2018.
18. Understanding alcohol use disorder. National Institute on Alcohol Abuse and Alcoholism. 2020. URL: <https://www.niaaa.nih.gov/alcohols-effects-health/alcohol-use-disorder> [accessed 2021-11-01]
19. Duko B, Ayalew M, Ayano G. The prevalence of alcohol use disorders among people living with HIV/AIDS: a systematic review and meta-analysis. *Subst Abuse Treat Prev Policy* 2019 Nov 14;14(1):52. [doi: [10.1186/s13011-019-0240-3](https://doi.org/10.1186/s13011-019-0240-3)]
20. Necho M, Belete A, Getachew Y. The prevalence and factors associated with alcohol use disorder among people living with HIV/AIDS in Africa: a systematic review and meta-analysis. *Subst Abuse Treat Prev Policy* 2020 Aug 24;15(1):63. [doi: [10.1186/s13011-020-00301-6](https://doi.org/10.1186/s13011-020-00301-6)]
21. O'Donnell A, Anderson P, Newbury-Birch D, Schulte B, Schmidt C, Reimer J, et al. The impact of brief alcohol interventions in primary healthcare: a systematic review of reviews. *Alcohol Alcohol* 2014 Nov 13;49(1):66-78 [FREE Full text] [doi: [10.1093/alcac/agt170](https://doi.org/10.1093/alcac/agt170)] [Medline: [24232177](https://pubmed.ncbi.nlm.nih.gov/24232177/)]
22. Murray E. Screening and brief intervention for alcohol use disorders in primary care. *BMJ* 2013 Jan 09;346(jan09 2):e8706. [doi: [10.1136/bmj.e8706](https://doi.org/10.1136/bmj.e8706)]
23. Kaner E, Beyer F, Muirhead C, Campbell F, Pienaar E, Bertholet N. Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database Syst Rev* 2018;2(2):cd004148. [doi: [10.1002/14651858.cd004148.pub4](https://doi.org/10.1002/14651858.cd004148.pub4)]
24. Angus C, Latimer N, Preston L, Li J, Purshouse R. What are the implications for policy makers? A systematic review of the cost-effectiveness of screening and brief interventions for alcohol misuse in primary care. *Front Psychiatry* 2014 Sep 01;5:114 [FREE Full text] [doi: [10.3389/fpsyt.2014.00114](https://doi.org/10.3389/fpsyt.2014.00114)] [Medline: [25225487](https://pubmed.ncbi.nlm.nih.gov/25225487/)]
25. Tariq L, van den Berg M, Hoogenveen RT, van Baal PH. Cost-effectiveness of an opportunistic screening programme and brief intervention for excessive alcohol use in primary care. *PLoS One* 2009 May 27;4(5):e5696 [FREE Full text] [doi: [10.1371/journal.pone.0005696](https://doi.org/10.1371/journal.pone.0005696)] [Medline: [19479081](https://pubmed.ncbi.nlm.nih.gov/19479081/)]
26. Kraemer KL. The cost-effectiveness and cost-benefit of screening and brief intervention for unhealthy alcohol use in medical settings. *Substance Abuse* 2007 Nov 21;28(3):67-77. [doi: [10.1300/j465v28n03\\_07](https://doi.org/10.1300/j465v28n03_07)]
27. Blackburn NA, Go VF, Bui Q, Hutton H, Tampi RP, Sripaipan T, et al. Implementation of two alcohol reduction interventions among persons with hazardous alcohol use who are living with HIV in Thai Nguyen, Vietnam: a micro-costing analysis. *Global Health Action* 2020 Sep 07;13(1):1814035. [doi: [10.1080/16549716.2020.1814035](https://doi.org/10.1080/16549716.2020.1814035)]
28. Nadkarni A, Weiss HA, Weobong B, McDaid D, Singla DR, Park A, et al. Sustained effectiveness and cost-effectiveness of Counselling for Alcohol Problems, a brief psychological treatment for harmful drinking in men, delivered by lay counsellors in primary care: 12-month follow-up of a randomised controlled trial. *PLoS Med* 2017 Sep 12;14(9):e1002386. [doi: [10.1371/journal.pmed.1002386](https://doi.org/10.1371/journal.pmed.1002386)]
29. Johnson NA, Kypri K, Latter J, Dunlop A, Brown A, Saitz R, et al. Effect of electronic brief intervention on uptake of specialty treatment in hospital outpatients with likely alcohol dependence: pilot randomized trial and qualitative interviews. *Drug Alcohol Dependence* 2018 Feb;183:253-260. [doi: [10.1016/j.drugalcdep.2017.11.016](https://doi.org/10.1016/j.drugalcdep.2017.11.016)]
30. Karno MP, Rawson R, Rogers B, Spear S, Grella C, Mooney LJ, et al. Effect of screening, brief intervention and referral to treatment for unhealthy alcohol and other drug use in mental health treatment settings: a randomized controlled trial. *Addiction* 2020 Jul 05;116(1):159-169. [doi: [10.1111/add.15114](https://doi.org/10.1111/add.15114)]
31. Borus J, Parhami I, Levy S. Screening, brief intervention, and referral to treatment. *Child Adolesc Psychiatr Clin N Am* 2016 Oct;25(4):579-601. [doi: [10.1016/j.chc.2016.05.001](https://doi.org/10.1016/j.chc.2016.05.001)] [Medline: [27613340](https://pubmed.ncbi.nlm.nih.gov/27613340/)]
32. Farley J, Miller E, Zamani A, Tepper V, Morris C, Oyegunle M, et al. Screening for hazardous alcohol use and depressive symptomatology among HIV-infected patients in Nigeria: prevalence, predictors, and association with adherence. *J Int Assoc Physicians AIDS Care (Chic)* 2010 Aug 26;9(4):218-226 [FREE Full text] [doi: [10.1177/1545109710371133](https://doi.org/10.1177/1545109710371133)] [Medline: [20798401](https://pubmed.ncbi.nlm.nih.gov/20798401/)]
33. Harris SK, Knight JR. Putting the screen in screening: technology-based alcohol screening and brief interventions in medical settings. *Alcohol Res* 2014;36(1):63-79 [FREE Full text] [Medline: [26259001](https://pubmed.ncbi.nlm.nih.gov/26259001/)]
34. Harris S, Louis-Jacques J, Knight J. Screening and brief intervention for alcohol and other abuse. *Adolesc Med State Art Rev* 2014;25(1):126-156. [doi: [10.1542/9781581108903-screening](https://doi.org/10.1542/9781581108903-screening)]
35. Johnson NA, Kypri K, Saunders JB, Saitz R, Attia J, Latter J, et al. Effect of electronic screening and brief intervention on hazardous or harmful drinking among adults in the hospital outpatient setting: a randomized, double-blind, controlled trial. *Drug Alcohol Depend* 2018 Oct 01;191:78-85. [doi: [10.1016/j.drugalcdep.2018.06.030](https://doi.org/10.1016/j.drugalcdep.2018.06.030)] [Medline: [30096637](https://pubmed.ncbi.nlm.nih.gov/30096637/)]
36. Jonas DE, Garbutt JC. Screening and counseling for unhealthy alcohol use in primary care settings. *Med Clin North Am* 2017 Jul;101(4):823-837. [doi: [10.1016/j.mcna.2017.03.011](https://doi.org/10.1016/j.mcna.2017.03.011)] [Medline: [28577629](https://pubmed.ncbi.nlm.nih.gov/28577629/)]

37. Kane JC, Murray LK, Bass JK, Johnson RM, Bolton P. Validation of a substance and alcohol use assessment instrument among orphans and vulnerable children in Zambia using Audio Computer Assisted Self-Interviewing (ACASI). *Drug Alcohol Dependence* 2016 Sep;166:85-92. [doi: [10.1016/j.drugalcdep.2016.06.026](https://doi.org/10.1016/j.drugalcdep.2016.06.026)]
38. Kilian Wells G, Moonie S, Pharr J. Association between alcohol screening and brief intervention during routine check-ups and alcohol consumption among adults living in California. *Archives Psychiatric Nursing* 2018 Dec;32(6):872-877. [doi: [10.1016/j.apnu.2018.07.001](https://doi.org/10.1016/j.apnu.2018.07.001)]
39. Kim T, Saitz R, Kretsch N, Cruz A, Winter M, Shanahan C. Screening for unhealthy alcohol and other drug use by health educators: do primary care clinicians document screening results? *J Addict Med. J Addict Med* 2013;7(3):204-209. [doi: [10.1097/adm.0b013e31828da017](https://doi.org/10.1097/adm.0b013e31828da017)]
40. Staton CA, Vissoci JR, Wojcik R, Hirshon JM, Mvungi M, Mmbaga BT, et al. Perceived barriers by health care providers for screening and management of excessive alcohol use in an emergency department of a low-income country. *Alcohol* 2018 Sep;71:65-73. [doi: [10.1016/j.alcohol.2018.01.003](https://doi.org/10.1016/j.alcohol.2018.01.003)]
41. Bajunirwe F, Haberer JE, Boum Y, Hunt P, Mocello R, Martin JN, et al. Comparison of self-reported alcohol consumption to phosphatidylethanol measurement among HIV-infected patients initiating antiretroviral treatment in southwestern Uganda. *PLoS One* 2014 Dec 1;9(12):e113152 [FREE Full text] [doi: [10.1371/journal.pone.0113152](https://doi.org/10.1371/journal.pone.0113152)] [Medline: [25436894](https://pubmed.ncbi.nlm.nih.gov/25436894/)]
42. Hahn J, Fatch R, Kabami J, Mayanja B, Emenyonu N, Martin J, et al. Self-report of alcohol use increases when specimens for alcohol biomarkers are collected in persons with HIV in Uganda. *J Acquir Immune Defic Syndr* 2012 Dec 01;61(4):e63-e64 [FREE Full text] [doi: [10.1097/QAI.0b013e318267c0f1](https://doi.org/10.1097/QAI.0b013e318267c0f1)] [Medline: [23138732](https://pubmed.ncbi.nlm.nih.gov/23138732/)]
43. Muyindike WR, Lloyd-Travaglini C, Fatch R, Emenyonu NI, Adong J, Ngabirano C, et al. Phosphatidylethanol confirmed alcohol use among ART-naïve HIV-infected persons who denied consumption in rural Uganda. *AIDS Care* 2017 Nov 13;29(11):1442-1447 [FREE Full text] [doi: [10.1080/09540121.2017.1290209](https://doi.org/10.1080/09540121.2017.1290209)] [Medline: [28278568](https://pubmed.ncbi.nlm.nih.gov/28278568/)]
44. Hahn JA, Bwana MB, Javors MA, Martin JN, Emenyonu NI, Bangsberg DR. Biomarker testing to estimate under-reported heavy alcohol consumption by persons with HIV initiating ART in Uganda. *AIDS Behav* 2010 Dec 10;14(6):1265-1268 [FREE Full text] [doi: [10.1007/s10461-010-9768-y](https://doi.org/10.1007/s10461-010-9768-y)] [Medline: [20697796](https://pubmed.ncbi.nlm.nih.gov/20697796/)]
45. Jain J, Evans J, Briceño A, Page K, Hahn J. Comparison of phosphatidylethanol results to self-reported alcohol consumption among young injection drug users. *Alcohol* 2014;49(5):520-524. [doi: [10.1093/alcalc/agu037](https://doi.org/10.1093/alcalc/agu037)]
46. Magidson JF, Fatch R, Orrell C, Amanyire G, Haberer JE, Hahn JA, META team. Biomarker-measured unhealthy alcohol use in relation to CD4 count among individuals starting ART in sub-Saharan Africa. *AIDS Behav* 2019 Jun 17;23(6):1656-1667 [FREE Full text] [doi: [10.1007/s10461-018-2364-2](https://doi.org/10.1007/s10461-018-2364-2)] [Medline: [30560484](https://pubmed.ncbi.nlm.nih.gov/30560484/)]
47. Bendtsen M. An electronic screening and brief intervention for hazardous and harmful drinking among Swedish university students: reanalysis of findings from a randomized controlled trial using a Bayesian framework. *J Med Internet Res* 2019 Dec 17;21(12):e14420 [FREE Full text] [doi: [10.2196/14420](https://doi.org/10.2196/14420)] [Medline: [31845903](https://pubmed.ncbi.nlm.nih.gov/31845903/)]
48. Bendtsen P, Bendtsen M, Karlsson N, White IR, McCambridge J. Online alcohol assessment and feedback for hazardous and harmful drinkers: findings from the AMADEUS-2 randomized controlled trial of routine practice in Swedish universities. *J Med Internet Res* 2015 Jul 09;17(7):e170 [FREE Full text] [doi: [10.2196/jmir.4020](https://doi.org/10.2196/jmir.4020)] [Medline: [26159179](https://pubmed.ncbi.nlm.nih.gov/26159179/)]
49. Bendtsen P, McCambridge J, Bendtsen M, Karlsson N, Nilsen P. Effectiveness of a proactive mail-based alcohol internet intervention for university students: dismantling the assessment and feedback components in a randomized controlled trial. *J Med Internet Res* 2012 Oct 31;14(5):e142 [FREE Full text] [doi: [10.2196/jmir.2062](https://doi.org/10.2196/jmir.2062)] [Medline: [23113955](https://pubmed.ncbi.nlm.nih.gov/23113955/)]
50. Bertholet N, Cunningham JA, Adam A, McNeely J, Daepfen J. Electronic screening and brief intervention for unhealthy alcohol use in primary care waiting rooms – A pilot project. *Substance Abuse* 2019 Jul 31;41(3):347-355. [doi: [10.1080/08897077.2019.1635963](https://doi.org/10.1080/08897077.2019.1635963)]
51. Bewick BM, Trusler K, Barkham M, Hill AJ, Cahill J, Mulhern B. The effectiveness of web-based interventions designed to decrease alcohol consumption — A systematic review. *Preventive Med* 2008 Jul;47(1):17-26. [doi: [10.1016/j.ypmed.2008.01.005](https://doi.org/10.1016/j.ypmed.2008.01.005)]
52. Browne FA, Wechsberg WM, Kizakevich PN, Zule WA, Bonner CP, Madison AN, et al. mHealth versus face-to-face: study protocol for a randomized trial to test a gender-focused intervention for young African American women at risk for HIV in North Carolina. *BMC Public Health* 2018 Aug 6;18(1):982. [doi: [10.1186/s12889-018-5796-8](https://doi.org/10.1186/s12889-018-5796-8)]
53. Chang G, Ondersma SJ, Blake-Lamb T, Gilstad-Hayden K, Orav EJ, Yonkers KA. Identification of substance use disorders among pregnant women: a comparison of screeners. *Drug Alcohol Dependence* 2019 Dec;205:107651. [doi: [10.1016/j.drugalcdep.2019.107651](https://doi.org/10.1016/j.drugalcdep.2019.107651)]
54. Crouce JM, Bittinger JN, Liu J, Kilmer JR. Electronic feedback in college student drinking prevention and intervention. *Alcohol Res* 2014;36(1):47-62 [FREE Full text] [Medline: [26259000](https://pubmed.ncbi.nlm.nih.gov/26259000/)]
55. Hallett J, Maycock B, Kypri K, Howat P, McManus A. Development of a web-based alcohol intervention for university students: processes and challenges. *Drug Alcohol Rev* 2009;28(1):31-39. [doi: [10.1111/j.1465-3362.2008.00008.x](https://doi.org/10.1111/j.1465-3362.2008.00008.x)]
56. Han BH, Masukawa K, Rosenbloom D, Kuerbis A, Helmuth E, Liao DH, et al. Use of web-based screening and brief intervention for unhealthy alcohol use by older adults. *J Substance Abuse Treatment* 2018 Mar;86:70-77. [doi: [10.1016/j.jsat.2018.01.002](https://doi.org/10.1016/j.jsat.2018.01.002)]

57. Han BH, Sherman SE, Link AR, Wang B, McNeely J. Comparison of the Substance Use Brief Screen (SUBS) to the AUDIT-C and ASSIST for detecting unhealthy alcohol and drug use in a population of hospitalized smokers. *J Substance Abuse Treatment* 2017 Aug;79:67-74. [doi: [10.1016/j.jsat.2017.05.014](https://doi.org/10.1016/j.jsat.2017.05.014)]
58. Paul CL, Carey M, Yoong SL, D'Este C, Makeham M, Henskens F. Access to chronic disease care in general practice: the acceptability of implementing systematic waiting-room screening using computer-based patient-reported risk status. *Br J Gen Pract* 2013 Sep 01;63(614):e620-e626. [doi: [10.3399/bjgp13x671605](https://doi.org/10.3399/bjgp13x671605)]
59. Rose GL, Skelly JM, Badger GJ, Maclean CD, Malgeri MP, Helzer JE. Automated screening for at-risk drinking in a primary care office using interactive voice response. *J Stud Alcohol Drugs* 2010 Sep;71(5):734-738. [doi: [10.15288/jsad.2010.71.734](https://doi.org/10.15288/jsad.2010.71.734)]
60. Goodyear-Smith F, Warren J, Bojic M, Chong A. eCHAT for lifestyle and mental health screening in primary care. *Ann Fam Med* 2013 Sep 09;11(5):460-466 [FREE Full text] [doi: [10.1370/afm.1512](https://doi.org/10.1370/afm.1512)] [Medline: [24019278](https://pubmed.ncbi.nlm.nih.gov/24019278/)]
61. Noble NE, Paul CL, Carey ML, Sanson-Fisher RW, Blunden SV, Stewart JM, et al. A cross-sectional survey assessing the acceptability and feasibility of self-report electronic data collection about health risks from patients attending an Aboriginal Community Controlled Health Service. *BMC Med Inform Decis Mak* 2014 Apr 16;14(1):34 [FREE Full text] [doi: [10.1186/1472-6947-14-34](https://doi.org/10.1186/1472-6947-14-34)] [Medline: [24739205](https://pubmed.ncbi.nlm.nih.gov/24739205/)]
62. Elley CR, Dawes D, Dawes M, Price M, Draper H, Goodyear-Smith F. Screening for lifestyle and mental health risk factors in the waiting room: feasibility study of the Case-finding Health Assessment Tool. *Can Fam Physician* 2014 Nov;60(11):e527-e534 [FREE Full text] [Medline: [25551137](https://pubmed.ncbi.nlm.nih.gov/25551137/)]
63. Phillips A, Gomez G, Boily M, Garnett G. A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries. *Int J Epidemiol* 2010;39(6):1541-1555. [doi: [10.1093/ije/dyq114](https://doi.org/10.1093/ije/dyq114)]
64. Group N. The feasibility of audio computer-assisted self-interviewing in international settings. *AIDS* 2007;21:S49-S58. [doi: [10.1097/01.aids.0000266457.11020.f0](https://doi.org/10.1097/01.aids.0000266457.11020.f0)]
65. Kauffman R, Kauffman R. Color-coded audio computer-assisted self-interview for low-literacy populations. *Epidemiology* 2011;22(1):132-133. [doi: [10.1097/ede.0b013e3181f6fc27](https://doi.org/10.1097/ede.0b013e3181f6fc27)]
66. Lor M, Bowers BJ. Feasibility of audio-computer-assisted self-interviewing with color-coding and helper assistance (ACASI-H) for Hmong older adults. *Res Nurs Health* 2017 Aug 21;40(4):360-371 [FREE Full text] [doi: [10.1002/nur.21796](https://doi.org/10.1002/nur.21796)] [Medline: [28431187](https://pubmed.ncbi.nlm.nih.gov/28431187/)]
67. UNESCO Institute for Statistics. The World Bank. URL: <https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=UG> [accessed 2021-11-01]
68. Seebregts CJ, Mamlin BW, Biondich PG, Fraser HS, Wolfe BA, Jazayeri D, et al. Human factors for capacity building. Lessons learned from the OpenMRS implementers network. *Yearb Med Inform* 2018 Mar 07;19(01):13-20. [doi: [10.1055/s-0038-1638681](https://doi.org/10.1055/s-0038-1638681)]
69. Hahn JA, Emenyonu NI, Fatch R, Muyindike WR, Kekiibina A, Carrico AW, et al. Declining and rebounding unhealthy alcohol consumption during the first year of HIV care in rural Uganda, using phosphatidylethanol to augment self-report. *Addiction* 2015 Nov 05;111(2):272-279. [doi: [10.1111/add.13173](https://doi.org/10.1111/add.13173)]
70. Hahn J, Cheng D, Emenyonu N, Lloyd-Travaglini C, Fatch R, Shade S, et al. Alcohol use and HIV disease progression in an antiretroviral naive cohort. *J Acquir Immune Defic Syndr* 2018 Apr 15;77(5):492-501 [FREE Full text] [doi: [10.1097/QAI.0000000000001624](https://doi.org/10.1097/QAI.0000000000001624)] [Medline: [29303844](https://pubmed.ncbi.nlm.nih.gov/29303844/)]
71. Bush K, Kivlahan DR, McDonnell 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](https://doi.org/10.1001/archinte.158.16.1789)] [Medline: [9738608](https://pubmed.ncbi.nlm.nih.gov/9738608/)]
72. Dedoose Version 9.0.17, Web Application for Managing, Analyzing, and Presenting Qualitative and Mixed Method Research Data. Los Angeles, CA: SocioCultural Research Consultants, LLC; 2021.
73. Watkins DC. Rapid and rigorous qualitative data analysis. *Int J Qual Method* 2017 Jun 08;16(1):160940691771213. [doi: [10.1177/1609406917712131](https://doi.org/10.1177/1609406917712131)]
74. Android becomes most widely used mobile OS for the first time in Africa. Techloy. 2014 Feb 04. URL: <http://techloy.com/2014/02/04/android-becomes-widely-used-mobile-os-africa/> [accessed 2015-02-10]
75. CommCare. Dimagi. URL: <https://www.dimagi.com/commcare/> [accessed 2015-02-10]
76. Guest M, MacQueen KM, Namey EE. Applied Thematic Analysis. Thousand Oaks, California: SAGE Publications; 2011.
77. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *Int J Qual Method* 2016 Nov 29;5(1):80-92. [doi: [10.1177/160940690600500107](https://doi.org/10.1177/160940690600500107)]
78. Feldman BJ, Fredericksen RJ, Crane PK, Safren SA, Mugavero MJ, Willig JH, et al. Evaluation of the Single-Item Self-Rating Adherence Scale for Use in Routine Clinical Care of People Living with HIV. *AIDS Behav* 2012 Oct 30;17(1):307-318. [doi: [10.1007/s10461-012-0326-7](https://doi.org/10.1007/s10461-012-0326-7)]
79. Monahan PO, Shacham E, Reece M, Kroenke K, Ong'or W, Omollo O, et al. Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya. *J Gen Intern Med* 2009 Feb 20;24(2):189-197 [FREE Full text] [doi: [10.1007/s11606-008-0846-z](https://doi.org/10.1007/s11606-008-0846-z)] [Medline: [19031037](https://pubmed.ncbi.nlm.nih.gov/19031037/)]

80. Daftary A, Hirsch-Moverman Y, Kassie GM, Melaku Z, Gadisa T, Saito S, et al. A qualitative evaluation of the acceptability of an interactive voice response system to enhance adherence to isoniazid preventive therapy among people living with HIV in Ethiopia. *AIDS Behav* 2017 Nov 24;21(11):3057-3067 [FREE Full text] [doi: [10.1007/s10461-016-1432-8](https://doi.org/10.1007/s10461-016-1432-8)] [Medline: [27221743](https://pubmed.ncbi.nlm.nih.gov/27221743/)]
81. Adam A, Schwartz RP, Wu L, Subramaniam G, Laska E, Sharma G, et al. Electronic self-administered screening for substance use in adult primary care patients: feasibility and acceptability of the tobacco, alcohol, prescription medication, and other substance use (myTAPS) screening tool. *Addict Sci Clin Pract* 2019 Oct 15;14(1):39 [FREE Full text] [doi: [10.1186/s13722-019-0167-z](https://doi.org/10.1186/s13722-019-0167-z)] [Medline: [31615549](https://pubmed.ncbi.nlm.nih.gov/31615549/)]
82. Saitz R. Web-based screening and brief intervention for the spectrum of alcohol problems\*1. *Prevent Med* 2004 Nov;39(5):969-975. [doi: [10.1016/j.ypmed.2004.04.011](https://doi.org/10.1016/j.ypmed.2004.04.011)]
83. Kypri K, Langley JD, Saunders JB, Cashell-Smith ML, Herbison P. Randomized controlled trial of web-based alcohol screening and brief intervention in primary care. *Arch Intern Med* 2008 Mar 10;168(5):530-536. [doi: [10.1001/archinternmed.2007.109](https://doi.org/10.1001/archinternmed.2007.109)] [Medline: [18332300](https://pubmed.ncbi.nlm.nih.gov/18332300/)]
84. Kypri K, Saunders JB, Gallagher SJ. Acceptability of various brief intervention approaches for hazardous drinking among university students. *Alcohol Alcohol* 2003 Nov 01;38(6):626-628. [doi: [10.1093/alcalc/agg121](https://doi.org/10.1093/alcalc/agg121)] [Medline: [14633653](https://pubmed.ncbi.nlm.nih.gov/14633653/)]
85. Toner P, Böhnke JR, Andersen P, McCambridge J. Alcohol screening and assessment measures for young people: a systematic review and meta-analysis of validation studies. *Drug Alcohol Depend* 2019 Sep 01;202:39-49 [FREE Full text] [doi: [10.1016/j.drugalcdep.2019.01.030](https://doi.org/10.1016/j.drugalcdep.2019.01.030)] [Medline: [31299552](https://pubmed.ncbi.nlm.nih.gov/31299552/)]
86. Satre DD, Anderson AN, Leibowitz AS, Levine-Hall T, Slome S, Flamm J, et al. Implementing electronic substance use disorder and depression and anxiety screening and behavioral interventions in primary care clinics serving people with HIV: protocol for the Promoting Access to Care Engagement (PACE) trial. *Contemp Clin Trials* 2019 Sep;84:105833 [FREE Full text] [doi: [10.1016/j.cct.2019.105833](https://doi.org/10.1016/j.cct.2019.105833)] [Medline: [31446142](https://pubmed.ncbi.nlm.nih.gov/31446142/)]
87. McNeely J, Kumar PC, Rieckmann T, Sedlander E, Farkas S, Chollak C, et al. Barriers and facilitators affecting the implementation of substance use screening in primary care clinics: a qualitative study of patients, providers, and staff. *Addict Sci Clin Pract* 2018 Apr 09;13(1):8 [FREE Full text] [doi: [10.1186/s13722-018-0110-8](https://doi.org/10.1186/s13722-018-0110-8)] [Medline: [29628018](https://pubmed.ncbi.nlm.nih.gov/29628018/)]
88. Tang AM, Hamunime N, Adams RA, Kanyinga G, Fischer-Walker C, Agolory S, et al. Introduction of an alcohol-related electronic screening and brief intervention (eSBI) program to reduce hazardous alcohol consumption in Namibia's antiretroviral treatment (ART) program. *AIDS Behav* 2019 Nov 23;23(11):3078-3092 [FREE Full text] [doi: [10.1007/s10461-019-02648-9](https://doi.org/10.1007/s10461-019-02648-9)] [Medline: [31444711](https://pubmed.ncbi.nlm.nih.gov/31444711/)]
89. Takahashi R, Wilunda C, Magutah K, Mwaura-Tenambergen W, Atwoli L, Perngparn U. Evaluation of alcohol screening and community-based brief interventions in rural western Kenya: a quasi-experimental study. *Alcohol Alcohol* 2018 Jan 01;53(1):121-128. [doi: [10.1093/alcalc/agg083](https://doi.org/10.1093/alcalc/agg083)] [Medline: [29087434](https://pubmed.ncbi.nlm.nih.gov/29087434/)]
90. Petersen Williams P, Petersen Z, Sorsdahl K, Mathews C, Everett-Murphy K, Parry CD. Screening and brief interventions for alcohol and other drug use among pregnant women attending midwife obstetric units in Cape Town, South Africa: a qualitative study of the views of health care professionals. *J Midwifery Womens Health* 2015 Jul 28;60(4):401-409. [doi: [10.1111/jmwh.12328](https://doi.org/10.1111/jmwh.12328)] [Medline: [26220766](https://pubmed.ncbi.nlm.nih.gov/26220766/)]
91. Pengpid S, Peltzer K, van der Heever H, Skaal L. Screening and brief interventions for hazardous and harmful alcohol use among university students in South Africa: results from a randomized controlled trial. *Int J Environ Res Public Health* 2013 May 21;10(5):2043-2057 [FREE Full text] [doi: [10.3390/ijerph10052043](https://doi.org/10.3390/ijerph10052043)] [Medline: [23698697](https://pubmed.ncbi.nlm.nih.gov/23698697/)]
92. Pengpid S, Peltzer K, Skaal L, Van der Heever H. Screening and brief interventions for hazardous and harmful alcohol use among hospital outpatients in South Africa: results from a randomized controlled trial. *BMC Public Health* 2013 Jul 11;13(1):644 [FREE Full text] [doi: [10.1186/1471-2458-13-644](https://doi.org/10.1186/1471-2458-13-644)] [Medline: [23844552](https://pubmed.ncbi.nlm.nih.gov/23844552/)]
93. Donoghue K, Patton R, Phillips T, Deluca P, Drummond C. The effectiveness of electronic screening and brief intervention for reducing levels of alcohol consumption: a systematic review and meta-analysis. *J Med Internet Res* 2014 Jun 02;16(6):e142 [FREE Full text] [doi: [10.2196/jmir.3193](https://doi.org/10.2196/jmir.3193)] [Medline: [24892426](https://pubmed.ncbi.nlm.nih.gov/24892426/)]
94. Saunders EC, Moore SK, Gardner T, Farkas S, Marsch LA, McLeman B, et al. Screening for substance use in rural primary care: a qualitative study of providers and patients. *J Gen Intern Med* 2019 Dec 14;34(12):2824-2832 [FREE Full text] [doi: [10.1007/s11606-019-05232-y](https://doi.org/10.1007/s11606-019-05232-y)] [Medline: [31414355](https://pubmed.ncbi.nlm.nih.gov/31414355/)]

## Abbreviations

- ACASI:** audio computer-assisted self-interviewing
- ARV:** antiretroviral
- FGD:** focus group discussion
- IDI:** in-depth interview
- ISS:** Immune Suppression Syndrome
- MRRH:** Mbarara Regional Referral Hospital
- RA:** research assistant
- SASH:** Self-Administered Digital Screener for Health

**SSA:** sub-Saharan Africa

*Edited by A Mavragani; submitted 16.11.21; peer-reviewed by G Liu, D Drabarek; comments to author 14.03.22; revised version received 10.06.22; accepted 21.07.22; published 01.09.22.*

*Please cite as:*

*Emenyonu N, Kekibiina A, Woolf-King S, Kyampire C, Fatch R, Dawson-Rose C, Muyindike W, Hahn J  
Digital Health Screening in People With HIV in Uganda to Increase Alcohol Use Reporting: Qualitative Study on the Development and Testing of the Self-administered Digital Screener for Health*

*JMIR Form Res 2022;6(9):e35015*

*URL: <https://formative.jmir.org/2022/9/e35015>*

*doi: [10.2196/35015](https://doi.org/10.2196/35015)*

*PMID: [36048519](https://pubmed.ncbi.nlm.nih.gov/36048519/)*

©Nneka Emenyonu, Allen Kekibiina, Sarah Woolf-King, Catherine Kyampire, Robin Fatch, Carol Dawson-Rose, Winnie Muyindike, Judith Hahn. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 01.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Integrating Social Determinants of Health With Tobacco Treatment for Individuals With Opioid Use Disorder: Feasibility and Acceptability Study of Delivery Through Text Messaging

Hasmeena Kathuria<sup>1</sup>, MD; Divya Shankar<sup>1</sup>, MD; Vinson Cobb<sup>1</sup>, BA; Julia Newman<sup>1</sup>, MD; Katia Bulekova<sup>2</sup>, MSc; Scott Wertz<sup>3</sup>; Belinda Borrelli<sup>4</sup>, PhD

<sup>1</sup>The Pulmonary Center, Boston University School of Medicine, Boston, MA, United States

<sup>2</sup>Boston University, Boston, MA, United States

<sup>3</sup>Agile Health, Lincolnshire, IL, United States

<sup>4</sup>Henry M Goldman School of Dental Medicine, Boston University, Boston, MA, United States

**Corresponding Author:**

Hasmeena Kathuria, MD

The Pulmonary Center

Boston University School of Medicine

72 E Concord St, R304

Boston, MA, 02118

United States

Phone: 1 6176384860

Email: [hasmeena@bu.edu](mailto:hasmeena@bu.edu)

## Abstract

**Background:** Individuals with opioid use disorder (OUD) have a high prevalence of smoking and frequently experience unmet social determinants of health (SDOH), which may be barriers to smoking cessation. Hospitalization is an opportunity to encourage smoking cessation. Unfortunately, many clinicians do not provide tobacco treatment to support the maintenance of cessation achieved during hospitalization. Interventions are required to support these high-risk individuals after hospital discharge.

**Objective:** This study aimed to test the feasibility and acceptability of a 28-day SMS text messaging program tailored to individuals with OUD, which provides smoking cessation support and addresses unmet SDOH needs.

**Methods:** From July to December 2019, we enrolled 25 individuals who were hospitalized with tobacco dependence and OUD at our large safety net hospital. The SMS text messaging program was initiated during hospitalization and continued for 28 days. Participants were enrolled in either the *ready to quit within 30 days* or the *not ready to quit within 30 days* program based on their readiness to quit. Automated SMS text messages were sent twice daily for 4 weeks. The topics included health and cost benefits of quitting, both general and opioid specific (16 messages); managing mood and stress (8 messages); motivation, coping strategies, and encouragement (18 messages); addressing medication misconceptions (5 messages); links to resources to address substance use (2 messages providing links to the Massachusetts Substance Use Helpline and Boston Medical Center resources), tobacco dependence (1 message providing a link to the Massachusetts Quitline), and unmet SDOH needs (6 messages assessing SDOH needs with links to resources if unmet SDOH needs were identified). Questionnaires and interviews were conducted at baseline and at 2 and 4 weeks after enrollment.

**Results:** The participants were 56% (14/25) female, 36% (9/25) African American, 92% (23/25) unemployed, and 96% (24/25) Medicaid insured. Approximately 84% (21/25) activated the program, and none of the participants unsubscribed. Approximately 57% (12/21) completed either the 2- or 4-week questionnaires. Program satisfaction was high (overall mean 6.7, SD 0.8, range 1-7). Many perceived that the SMS text messaging program provided social support, companionship, and motivation to stop smoking. Messages about the health benefits of quitting were well received, whereas messages on how quitting cigarettes may prevent relapse from other substances had mixed views, highlighting the importance of tailoring interventions to patient preferences.

**Conclusions:** SMS text messaging to promote smoking cessation and address SDOH needs may be an effective tool for improving quit rates and health outcomes in individuals with tobacco dependence and OUD. Our study adds to the growing body of evidence that SMS text messaging approaches are feasible and acceptable for providing tobacco treatment to all individuals who smoke, even among low-income populations who have OUD and are not ready to quit.

**KEYWORDS**

text message; smoking cessation; opioid use disorder; tobacco dependence; tobacco treatment interventions; mobile phone

## Introduction

Individuals with opioid use disorder (OUD) smoke at rates as high as 83% to 97% [1,2]. Co-occurring tobacco and opioid use leads to high morbidity and mortality [3,4]. Smoking cessation improves tobacco health-related outcomes and could increase long-term abstinence from opioids [5,6]. A meta-analysis of 24 studies showed a positive impact of smoking cessation on substance use disorder outcomes; both tobacco treatment and smoking cessation either reduced or had no effect on other drug use [7].

Although individuals with OUD desire assistance with smoking cessation [8], tobacco treatment is infrequently offered [9-11]. Hospitalization is an opportunity to encourage smoking cessation [12-16]. We previously found a high acceptance of inpatient tobacco counseling among individuals who were hospitalized with OUD [17]. Unfortunately, many clinicians do not provide tobacco treatment to support the maintenance of cessation achieved during hospitalization [18]. Interventions are required to support these high-risk individuals after hospital discharge.

Individuals who smoke are 1.5 to 2 times more likely to quit smoking when enrolled in SMS text messaging programs for smoking cessation [19-21]. SMS text messaging is highly disseminative: mobile phone ownership is near universal; SMS text messaging is highly prevalent across race, education, and income, and >85% of individuals who smoke send and receive SMS text messages regularly [22-24]. Although studies have not assessed the efficacy of SMS text messaging for smoking cessation in individuals with OUD, they have shown a moderately high reach in Medicaid populations [25].

There is nearly universal agreement in scientific and public health communities that social determinants of health (SDOH)—the social circumstances in which people are born, grow, live, work, and age—influence access to resources and opportunities that affect health [26,27]. SDOH has a far greater impact on health outcomes than medical interventions [28]. For example, cross-sectional and longitudinal studies show that food insecurity is independently associated with an increased likelihood of smoking cigarettes, likely resulting from a combination of physiological factors, including stress, anxiety, and depression [29-31]. Thus, screening for and addressing unmet SDOH needs through policies and interventions may mitigate these factors and improve smoking cessation outcomes.

Many individuals with OUD also experience unmet SDOH needs (eg, transportation issues and food and housing insecurity) [32,33], which may be barriers to smoking cessation. However, studies have not systematically screened for unmet SDOH needs or provided referrals to address these needs in this population. Given their high smoking rates, tobacco-related comorbidities, lack of access to treatment, and inclusion in tobacco treatment trials, an integrative intervention combining tobacco treatment with SDOH assessment and referral may improve smoking

cessation among patients with OUD. We sought to iteratively develop and deploy an SMS text messaging program tailored to those with OUD, which provides smoking cessation support and resources to address SDOH. We report the results of a pilot feasibility and acceptability study of an SMS text messaging program initiated during hospitalization and continued for 28 days.

## Methods

### Recruitment and Enrollment

From July to December 2019, we enrolled 25 individuals who were hospitalized with tobacco dependence and OUD at the Boston Medical Center (BMC), the largest safety net hospital in New England. We identified participants from a list of individuals who were hospitalized, who triggered consultation with the Tobacco Treatment Consult service based on current smoking status in the electronic health record [34], and who had an International Classification of Diseases, 10th Revision diagnosis of OUD by chart review. Eligible participants were (1) aged  $\geq 18$  years, (2) hospitalized at the BMC, (3) able to speak and read English, (4) currently smoking cigarettes, (5) diagnosed with OUD, (6) mobile phone owners with an unlimited SMS text messaging plan, (7) in agreement to receive SMS text messages for 1 month, (8) not participating in other SMS text messaging or tobacco treatment programs, (9) able to receive a test SMS text message, and (10) able to provide informed consent. The participants were excluded if they were cognitively impaired.

A total of 96 participants met the screening criteria by electronic health record review (individuals listed as current for smoking status and OUD), of whom 42 (44%) were ineligible by face-to-face screening ( $n=34$ , 81% did not have unlimited SMS text messaging or had no phone at the time of hospitalization;  $n=3$ , 7% had stopped smoking;  $n=1$ , 2% did not have OUD; and  $n=4$ , 10% could not provide consent), 2 (2%) were unavailable, and 27 (28%) declined or were not interested in learning about the study.

### Ethics Approval

Following the provision of study information, 25 individuals agreed, provided informed consent, and were enrolled. Participants were compensated up to US \$50 for participation: US \$10 for completing the baseline survey and interview, US \$15 for completing a 2-week follow-up survey and interview, and US \$25 for completing a 4-week follow-up survey and interview. This study was approved by our institutional review board (protocol number H-38709).

### Structure of Program

The SMS text messaging program lasted 28 days, with the first day of SMS text messages sent during hospitalization. Participants were sent 2 intervention SMS text messages daily (9 AM and 5 PM) in addition to weekly SMS text message

assessments (see the *Measures* section). There were 2 tracks: one for individuals *ready to quit within 30 days* and the other for individuals *not ready to quit within 30 days*, as assessed by their answer to an introductory SMS text message assessment. The program had bidirectional or 2-way SMS text messaging capabilities; for example, participants could text a keyword (eg, *CRAVE*) to receive strategies and tips.

The SMS text messages were fully automated. All incoming SMS text messages were monitored and, if needed, responded to in real time by a team member via a password-protected dashboard interface if the system did not recognize an SMS text message and could not produce an automated response. SMS text messages were delivered by Agile Health, Inc, and their system is Health Insurance Portability and Accountability Act compliant. All the data managed by Agile Health and its message delivery partners were encrypted in transit and at rest. All user interactions, comprising solicited and unsolicited SMS text messages were recorded, including “STOP” (the standard keyword for unsubscribing). Agile Health, Inc notified the research team members of the urgent unsolicited SMS text messages sent to the server. All SMS text messages sent by the participant to, and responded by, the system were reviewed by the research team weekly.

### Program Content

The topics of the SMS text messages included (1) health and cost benefits of quitting, both general and opioid specific (16 messages); (2) managing mood and stress (8 messages); (3) motivation, coping strategies, and encouragement (18 messages); (4) addressing medication misconceptions (5 messages); (5) links to resources to address substance use (2 messages providing links to the Massachusetts Substance Use Helpline and BMC resources), tobacco dependence (1 message providing link to Massachusetts Quitline), and resources for unmet SDOH needs (6 messages assessing SDOH needs with links to resources if SDOH needs were unmet). Messages were obtained from three sources: (1) the National Cancer Institute’s Smokefree TXT [35], (2) content adapted from prior work by Borrelli et al [36], and (3) novel messages developed by the study team.

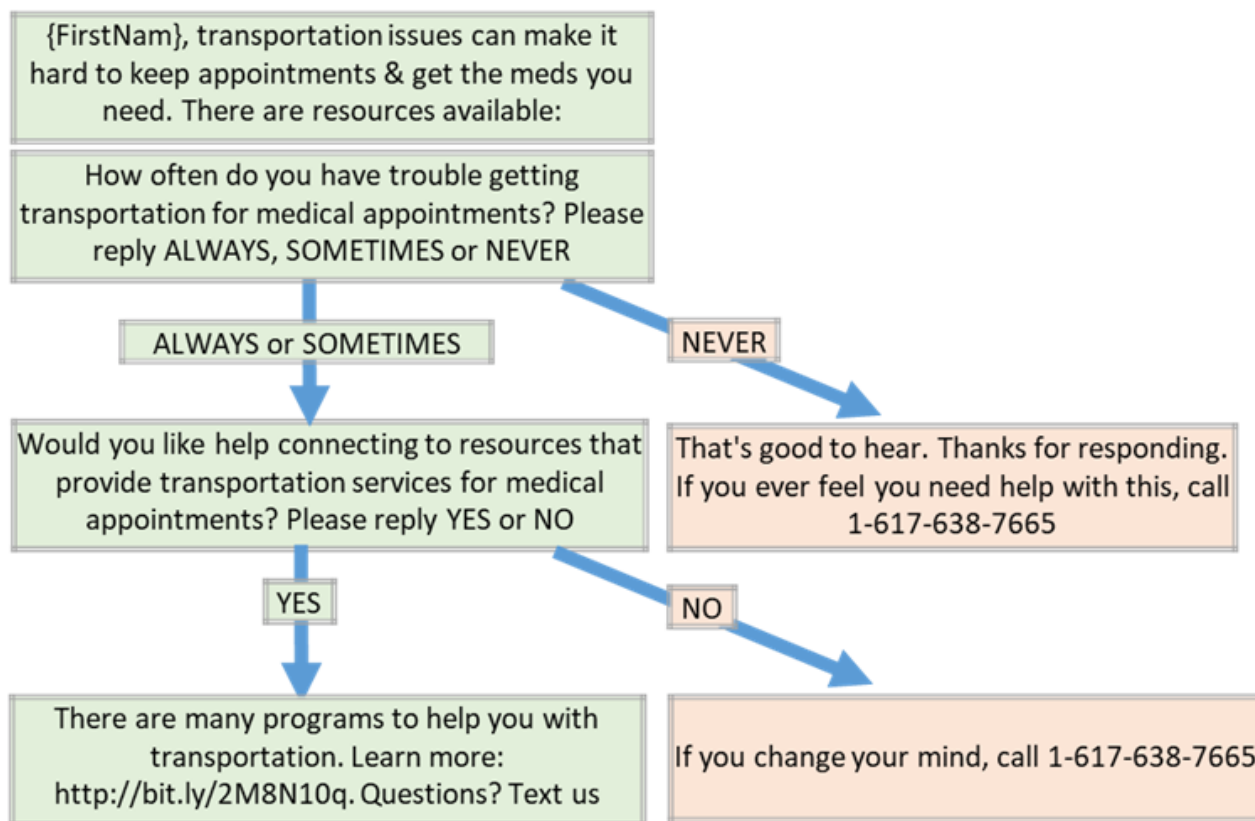
Assessments and links to resources were provided for the following 6 SDOH needs: difficulty with transportation to medical appointments, inability to pay for medications, risk of becoming homeless, food insecurity, trouble paying for heat or electricity, and the likelihood of needing to look for a job ([Multimedia Appendix 1](#)). Individuals were considered to have

an unmet SDOH need if they answered either “ALWAYS” or “SOMETIMES” to an SMS text message assessing for the SDOH need (example SMS text message: “How often do you have trouble getting transportation for medical appointments? Please reply ALWAYS, SOMETIMES, or NEVER”). Participants were then sent a link to resources if they answered “YES” to a text assessing their desire for help with that need (eg, “Would you like help connecting to resources that provide transportation services for medical appointments? Please reply YES or NO”). [Figure 1](#) outlines the algorithm for the provision of resources.

Assessments for the 6 SDOH needs were adapted from the validated Tool for Health and Resilience In Vulnerable Environments (THRIVE) screening tool and accompanying referral guide [37]; the THRIVE screening tool asks about 8 SDOH domains (housing, food, affording medications, transportation, utilities, caregiving, education, and employment) selected based on their impact on health and available services in the community. The THRIVE referral guide is a web-based directory of resources with contact information for community services to meet the SDOH needs. The need for caregiving and education were not assessed to reduce participant burden and also as we believed that addressing these SDOH needs would require a more nuanced discussion with an advocate or community health worker.

Several SMS text messages (3-5 messages per week) were customized based on the readiness to quit. In the *ready to quit* track, an SMS text message on coping strategies was “Cravings will get weaker and less frequent with every day that you don’t smoke.” In the *not ready to quit* track, a parallel message was “Don’t let cravings get in the way of deciding to quit. There are good meds to help with cravings. Cravings get weaker with each passing day.” Messages in the *not ready to quit* track were directed toward developing participants’ personal reasons for change and increasing their motivation and self-efficacy to stop smoking. Examples of such messages are “Thought: What is the best result you can imagine if you quit smoking? Imagine all the ways your life would change. How would you spend the extra money? How would you feel? Who would you spend time with? Where would you spend your time?” and “Thought for the day: Fall down 7 times, get up 8. The key to success is to persist even if you have previously failed.” Messages in the *ready to quit* track provided encouragement; an example was “Stay positive. Your journey to a smokefree life may be a struggle, but looking back it will be well worth it.”

**Figure 1.** Example algorithm for assessment of unmet social determinants of health needs and for provision of resources to address the need.



**Measures**

The study staff administered baseline questionnaires and interviews in person at the time of hospitalization and conducted 2- and 4-week questionnaires and interviews by telephone. Questionnaires followed by qualitative interviews were administered at the same encounter.

**Participant Characteristics**

Baseline demographics, use of substances, and comorbid mental health disorders were collected at the time of enrollment. Mobile technology use and smoking characteristics were assessed, including the level of cigarette dependence measured using the Fagerstrom Test for Cigarette Dependence [38,39]. Scores range from 0 to 10, with higher scores indicating more intense physical dependence on nicotine. No dependence corresponds to a score of 0, low dependence to a score of 1 or 2, low to moderate dependence to a score of 3 or 4, moderate dependence to a score of 5 to 7, and high dependence to a score of 8 to 10. We assessed psychosocial characteristics by assessing responses to the following question: “Have you ever been diagnosed with any of the following (check all that apply)? Answer choices: Depression; Anxiety; Bipolar Disorder; Manic-Depressive disorder; None of the above; Prefer not to answer.”

**Program Engagement and Interactions With Program**

Engagement was assessed through participants’ interactions with the program. We calculated the *totalresponse rate* by computing the number of participant-submitted responses to SMS text messages in which a response was expected and dividing this by the number of solicited responses. The number of *unsolicited* SMS text messages (messages sent by users where

a response was not expected, such as “Thanks you guys are a big help”) was an additional index of engagement.

**Program Satisfaction**

Program satisfaction was measured via 2- and 4-week questionnaires, using several indices of satisfaction. The “share-worthiness” of the SMS text messages was assessed by asking whether participants showed the SMS text messages to others (response: yes/no) and the extent to which they believed the SMS text messages would be helpful to family and friends (range 1=not at all helpful to 7=very helpful). The perceived quality of the SMS text messages was measured using 2 items from the Mobile Application Rating Scale [40]: one using star ratings (1 star=one of the worst SMS text message programs to 5 stars=one of the best SMS text message programs) and the other assessing how much longer they would have liked to receive the SMS text messages. Satisfaction with the program components and overall program satisfaction were assessed using 9 items (range 1=not satisfied at all to 7=very much satisfied). The likeability of program components was assessed using 7 items (range 1=did not like it at all to 7=liked it very much). The full list of items used to assess program satisfaction is presented in the *Results* section.

Satisfaction was additionally measured by eliciting responses on the helpfulness of the content of 8 specific SMS text messages (2 times per week, an intervention message was followed by an assessment SMS text message asking participants the following: “How helpful did you find this text? 3=Very helpful, 2=Neutral, or 1=Not helpful”). The 8 specific SMS text messages that were assessed for the helpfulness of the content are presented in the *Results* section.

### ***Perceived Impact of the Program***

Weekly SMS text messages assessed the degree to which the program was helpful in motivating smoking cessation (range 1=not helpful to 5=very helpful) [40]. The 2- and 4-week questionnaires assessed the perceived impact of the program on (1) motivation to stop smoking, (2) belief that stopping opioids and smoking cessation can occur simultaneously, and (3) knowledge of health risks from smoking (range 1=not at all to 7=very much).

### ***Qualitative Assessment of Feasibility, Acceptability, and Satisfaction***

We qualitatively measured feasibility and acceptability at 2 and 4 weeks using semistructured interview guides. The guide assessed the participants' (1) perceived impact in motivating smoking cessation, (2) experiences with the program, (3) preference of content, and (4) suggestions for improvement.

### **Data Analyses**

Basic descriptive statistics were calculated using SPSS (version 18; IBM Corp) to summarize the quantitative responses. For qualitative interviews, we used inductive content analysis to analyze transcripts and performed unstructured coding of the

transcripts to identify themes. A total of 2 members developed a codebook, independently reviewed all transcripts, and added codes until the team reached a consensus. We finalized the conceptual categories, grouped themes in each category, and identified quotes that best highlighted the themes. The interviews were audio recorded and transcribed verbatim.

## **Results**

### **Quantitative Analyses**

#### ***Participant Characteristics***

The participants were 56% (14/25) female, 36% (9/25) African American, 20% (5/25) Hispanic, 92% (23/25) unemployed, and 96% (24/25) Medicaid insured. The mean age was 45.8 (SD 11, range 28-63) years. Of the 25 participants, 12 (48%) reported current opioid use, and 12 (48%) were receiving medication-assisted treatment for OUD. Participants smoked for an average of 26.2 (SD 12, range 3-48) years. The Fagerstrom Test for Cigarette Dependence scores averaged 5.1 (SD 2), suggesting moderate nicotine dependence (Table 1). Approximately 44% (11/25) of the participants were ready to stop smoking within 30 days (Table 1).

**Table 1.** Characteristics of participants (N=25).

Baseline characteristics	All participants	Ready to quit track (n=14)	Not ready to quit track (n=11)
Age (years), mean (SD)	45.8 (11)	50.8 (9.5)	39.6 (8.8)
Sex (female), n (%)	14 (56)	7 (50)	7 (64)
<b>Race, n (%)</b>			
Black or African American	9 (36)	4 (29)	5 (45)
White	13 (52)	8 (57)	5 (45)
Other	3 (12)	2 (14)	1 (9)
Hispanic ethnicity, n (%)	5 (20)	3 (21)	2 (18)
Medicaid insurance, n (%)	24 (96)	13 (93)	11 (100)
<b>Education, n (%)</b>			
Less than high school	7 (28)	4 (29)	3 (27)
High school or General Educational Development	11 (44)	6 (43)	5 (45)
More than high school	7 (28)	4 (29)	3 (27)
<b>Housing situation, n (%)</b>			
At risk of homelessness	3 (12)	2 (14)	1 (9)
Experiencing homelessness	12 (48)	5 (36)	7 (64)
Divorced or separated, widowed, or never married, n (%)	20 (80)	12 (86)	8 (73)
Unemployed, n (%)	23 (92)	13 (93)	10 (91)
<b>Yearly household income before taxes (US \$), n (%)</b>			
0-14,999	17 (68)	9 (64)	8 (73)
>15,000	4 (16)	3 (21)	1 (9)
Prefer not to answer or do not know	4 (16)	2 (14)	2 (18)
Depression or anxiety, n (%)	19 (76)	9 (64)	10 (91)
<b>Current use of substances, n (%)</b>			
Alcohol ( $\geq 5$ for men and $\geq 4$ for women in 1 day)	5 (25)	3 (21)	1 (9)
Cocaine	13 (52)	4 (29)	9 (82)
Opiates	12 (48)	7 (50)	5 (45)
Marijuana	12 (48)	7 (50)	5 (45)
Prescription drugs (not prescribed)	6 (24)	2 (14)	4 (36)
Methamphetamines	4 (16)	1 (7)	3 (27)
<b>Smoking characteristics</b>			
Years smoked, mean (SD)	26.2 (12)	27.9 (13.4)	24.1 (9.4)
Smokes daily, n (%)	25 (100)	14 (100)	11 (100)
<b>Importance of quitting smoking, n (%)</b>			
Very important or important	17 (68)	12 (86)	5 (45)
Neutral	4 (16)	1 (7)	3 (27)
Low importance or not important	4 (16)	1 (7)	3 (27)
<b>Motivation to quit smoking, n (%)</b>			
Very motivated or motivated	16 (64)	12 (86)	4 (36)
Somewhat or slightly motivated	7 (28)	2 (14)	5 (45)
Not at all motivated	2 (8)	0 (0)	2 (18)
Fagerstrom score, mean (SD)	5.1 (2)	5.6 (2.4)	4.6 (1.4)
Dual use of cigarettes and e-cigarettes, n (%)	5 (25)	0 (0)	5 (45)

Baseline characteristics	All participants	Ready to quit track (n=14)	Not ready to quit track (n=11)
<b>Mobile technology, n (%)</b>			
Smartphone ownership	23 (92)	13 (93)	10 (91)
<b>SMS text messages sent per day per week, n (%)</b>			
2-9 per day	7 (28)	5 (36)	2 (18)
>10 per day	17 (68)	9 (64)	8 (73)
2-6 per week	1 (4)	0 (0)	1 (9)

### **Engagement and Interactions With the Text Messaging Program**

While in the hospital, and after obtaining informed consent, participants were asked via SMS text message whether they would like to begin (*activate*) the program. Of the 25 participants, 21 (84%; n=11, 52% *ready to quit*, and n=10, 48% *not ready to quit*) responded with *yes*. Of the 25 participants, 2 (8%) did not activate the program and expressed concern about not having access to a phone after discharge (eg, attending a rehabilitation program and concern about minutes); the other 2 (8%) were not available to assess reasons for not activating. Of the 21 participants, 13 (62%; n=5, 38% *ready to quit*, and n=8, 62% *not ready to quit*) submitted at least one response to the helpfulness assessments of 8 intervention SMS text messages (55/168, 32.7% response rate); 13 (62%; n=6, 46% *ready to quit*, and n=7, 54% *not ready to quit*) submitted at least one response to the SMS text messages assessing SDOH needs (44/147, 29.9% response rate); 15 (71%; n=7, 47% *ready to quit*, and n=8, 53% *not ready to quit*) responded to at least one weekly text SMS message assessment regarding the degree to which the program was helpful in motivating cessation (35/84, 42% response rate); and 14 (67%; n=7, 50% *ready to quit*, n=8, 57% *not ready to quit*) sent at least one unsolicited SMS text message for a total of 143 unsolicited messages. None of the participants unsubscribed.

### **Perceptions of Program**

Of the 21 participants, 12 (57%; n=7, 58% *ready to quit*, and n=5, 42% *not ready to quit*) completed either the 2- or 4-week questionnaires: 6 (50%; n=3, 50% *ready to quit*, and n=3, 50% *not ready to quit*) completed both questionnaires (2- and 4-week

responses were similar; only 4-week responses were analyzed); 4 (33%; n=2, 50% *ready to quit*, and n=2, 50% *not ready to quit*) completed only the 2-week questionnaire; and 2 (17%; n=1, 50% *ready to quit*, and n=1, 50% *not ready to quit*) completed only the 4-week questionnaire.

### **Program Satisfaction**

Of the 12 participants, 10 (83%; n=6, 60% *ready to quit*, and n=4, 40% *not ready to quit*) rated the program  $\geq 4$  stars, and 2 (16%; n=1, 50% *ready to quit*, and n=1, 50% *not ready to quit*) rated the program 3 stars. Of the 12 participants, 4 (33%; n=3, 75% *ready to quit*, and n=1, 25% *not ready to quit*) wanted the program to last up to 2 months longer, and 8 (67%; n=4, 50% *ready to quit*, n=4, 50% *not ready to quit*) indicated that they wanted the program to last  $\geq 3$  months longer. None of the participants thought that the program had interfered with their schedules. Of the 12 participants, 9 (75%; n=5, 56% *ready to quit*, and n=4, 44% *not ready to quit*) shared the SMS text messages with others. Participants believed that the SMS text messages would be helpful to family and friends (mean 5.4, SD 1.1, range 1-7; *ready to quit* mean 5.1, SD 1; *not ready to quit* mean 5.6, SD 1). Of the 12 participants, 11 (92%; n=7, 64% *ready to quit*, and n=4, 36% *not ready to quit*) were very likely or likely to recommend the program to others, and 1 (8%) person in the *not ready to quit* track was somewhat likely to recommend the program to others.

Overall, program satisfaction was high (overall mean 6.7, SD 0.8, range 1-7; *ready to quit* mean 7, SD 0; *not ready to quit* mean 6.2, SD 0.98). Participants reported that the content was trustworthy (mean 6.5, SD 0.8). Most liked the frequency (mean 5.8, SD 1.8, range 1-7) and timing (mean 5.9, SD 1.2, range 1-7) of the SMS text messages (Table 2).

**Table 2.** Questionnaire responses on likability, satisfaction, and perceived impact of the SMS text messaging program (N=12).

Items	Overall, mean (SD)	<i>Ready to quit</i> track (n=7), mean (SD)	<i>Not ready to quit</i> track (n=5), mean (SD)
<b>Likeability scale items<sup>a</sup></b>			
Having to respond to SMS text message questions	5.4 (2.2)	5 (2.6)	6 (1.3)
The degree to which the program was interesting	5.3 (1)	5.1 (1)	5.6 (1)
The degree to which the program was useful	5.8 (1.3)	5.7 (1.2)	6 (1.5)
The degree to which the program was engaging	5.7 (1.4)	6 (1.3)	5.2 (1.3)
The degree to which the program was boring	1.8 (1.1)	1.9 (1.4)	1.8 (0.7)
The frequency with which texts were delivered	5.8 (1.7)	5.9 (2.1)	5.8 (1.2)
The time of the day that texts were received	5.9 (1.2)	6.3 (1.2)	5.4 (1)
<b>Program satisfaction scale items<sup>b</sup></b>			
Overall satisfaction with the program	6.7 (0.8)	7 (0)	6.2 (1)
Receipt of support when needed	5.9 (1.1)	5.6 (1.2)	6.4 (0.8)
The amount of information in the SMS text messages	5.9 (1.2)	6.1 (1.1)	5.6 (1.2)
The quality of the information in the SMS text messages	5.8 (1.3)	5.9 (1.4)	5.6 (1.2)
Relevancy of program for self	5.4 (1.8)	5.4 (2.1)	5.4 (1.5)
The trustworthiness of the information	6.5 (0.8)	6.7 (0.7)	6.2 (0.75)
The level of program customization	5.6 (1.3)	5.6 (1.2)	5.6 (1.5)
The degree to which the SMS text messages were well written	6.8 (0.6)	6.9 (0.3)	6.6 (0.8)
The degree to which the SMS text messages were easy to integrate into routine	6.2 (1.1)	6.1 (1.4)	6.2 (0.7)
<b>Perceived impact of the program on motivation to stop smoking, beliefs, and overall knowledge of health risk<sup>c</sup></b>			
The degree to which the program motivated you to quit smoking	6.2 (1.2)	6.7 (0.5)	5.4 (1)
Belief that opioids and smoking cessation can occur at the same time	4.5 (2.5)	4.7 (2.7)	4.4 (2.3)
Overall knowledge about the health risk of smoking	5.6 (1.9)	6.1 (1.4)	4.8 (2.2)

<sup>a</sup>Range: 1=did not like it at all to 7=liked it very much.

<sup>b</sup>Range: 1=not satisfied at all to 7=very much satisfied.

<sup>c</sup>Range: 1=not helpful to 7=very helpful in motivating smoking cessation.

### Perception of Program Content

Of the 12 participants, 9 (75%; n=5, 56% *ready to quit*, and n=4, 44% *not ready to quit*) rated SDOH the SMS text messages (eg, where to find food pantries) *helpful*, and 10 (83%; n=6, 60% *ready to quit*, and n=4, 40% *not ready to quit*) rated the SMS text messages on managing mood and stress as *helpful*. All 12 participants rated the SMS text messages on resources for quitting smoking as *helpful*, and 11 (92%; n=6, 55% *ready to quit*, and n=5, 45% *not ready to quit*) rated the SMS text

messages about where to find help for other substances as *helpful*. The overall response rates for the SMS text message assessments on helpfulness (3=very helpful, 2=neutral, and 1=not helpful) of the 8 specific intervention messages were low (response rates ranged from 23.8% to 57.1%; [Table 3](#)). Participants in both tracks gave high ratings for messages about managing mood and stress, addressing medication misconceptions, and increasing their motivation to quit. Participants were neutral regarding messages about the benefits of quitting smoking on the use of other substances ([Table 3](#)).



**Table 3.** Response rates and ratings of 8 specific SMS text messages that were assessed for the helpfulness of content (N=21)<sup>a</sup>.

Rating	Text (specific SMS text messages assessed)	Total		Ready to quit track		Not ready to quit track	
		Response rate (%)	Rating, mean (SD)	Response rate (n=11; %)	Rating, mean (SD)	Response rate (n=10; %)	Rating, mean (SD)
Helpful 1: managing mood and stress	“Stress Tip: Talk about your problems! This lowers stress and gives new perspectives. Holding it in could affect your health and wellness”	57.1	2.7 (0.5)	45.5	2.6 (0.5)	70	2.7 (0.5)
Helpful 2: tips for cravings	“Whenever you want a cig, try the four D’s: Delay, Deep breathe, Drink water, Do something to take your mind off smoking”	28.6	2.5 (0.5)	36.3	2.5 (0.5)	20	2.5 (0.5)
Helpful 3: addressing medication mistrust	“MYTH: Chantix/Wellbutrin will make me feel depressed. FACT: Research shows no evidence that these meds increase risk of suicide & depression”	28.7	2.7 (0.5)	18.2	2.5 (0.5)	40	2.7 (0.5)
Helpful 4: motivating to quit	“Thought for the day: What would get easier in your life if you didn’t smoke? No more worrying about finding money to buy cigarettes and where you can smoke. Less worry about your health. What else would get better?”	42.9	2.6 (0.7)	36.3	3 (0)	50	2.2 (0.7)
Helpful 5: benefits of quitting (opiate specific)	“Smokers in substance use treatment are more likely to die from smoking-related disease compared to complications of their current drug use”	38.1	2.1 (0.8)	36.3	2.5 (0.9)	40	1.8 (0.4)
Helpful 6: motivating to quit	“It might seem like you are giving up a lot when you stop smoking, try to think about all you are gaining”	28.6	2.8 (0.4)	27.3	3 (0)	30	2.7 (0.5)
Helpful 7: benefits of quitting (opiate specific)	“MYTH: Quitting cigarettes could negatively impact recovery. FACT: Smoking cessation may promote recovery in patients who use opioids”	23.8	2.4 (0.5)	27.3	2.7 (0.5)	20	2 (0)
Helpful 8: managing mood and stress	“What pleasure do you get from smoking? Find healthier alternatives in your life that can bring you these same feelings”	33.3	2.9 (0.3)	45.5	2.8 (0.4)	20	3 (0)

<sup>a</sup>Ratings: 3=very helpful, 2=neutral, and 1=not helpful.

### Response Rates of Assessments on SDOH Needs

The response rates for SMS text messages assessing SDOH needs varied. All 13 individuals who responded had at least one unmet SDOH need: 5 (38%) had 1 unmet need, 3 (23%) had 2 unmet needs, 1 (8%) had 3 unmet needs, and 4 (31%) had 4 unmet needs. Responses for SDOH needs were as follows: trouble getting transportation for medical appointments (8/13, 62%; n=5, 63% always; n=2, 25% sometimes; and n=1, 13% never), trouble paying for medications (10/13, 77%; n=1, 10% always; n=4, 40% sometimes; and n=5, 50% never), risk of becoming homeless (5/13, 38%; n=3, 60% high; n=1, 20% medium; and n=1, 20% low), frequency of running out of food without having money to pay for more (4/13, 31%; n=2, 50% often; n=2, 50% sometimes; and n=0, 0% never), trouble paying for heat or electricity (5/13, 38%; n=1, 20% always; n=2, 40% sometimes; and n=2, 40% never), and likelihood of looking for a job in the near future (6/13, 46%; n=2, 33% high; n=3, 50% low; and n=1, 17% none).

### Perceived Impact of Program on Motivation to Stop Smoking

All 12 participants who completed the questionnaires agreed or strongly agreed that participating in the program made them think about quitting smoking. Of the 12 participants, 8 (67%; n=4, 50% *ready to quit*, and n=4, 50% *not ready to quit*) agreed or strongly agreed with “The program made me think that it is okay to quit tobacco and other drugs at the same time”; 1 (8%) participant in the *ready to quit* track was undecided, and 3 (38%; n=2, 67% *ready to quit*, and n=1, 33% *not ready to quit*) disagreed. Participants perceived that the program increased their motivation to quit (overall mean 6.2, SD 1, range 1-7; *ready to quit* mean 6.7, SD 0.5; *not ready to quit* mean 5.4, SD 1; Table 2).

Of the 21 participants, 15 (71%) individuals who responded to at least one weekly SMS text message assessment on the helpfulness in motivating smoking cessation (1=not helpful to 5=very helpful) believed the program was helpful: 12 (response

rate 57%; mean 3.6, SD 0.85) in week 1, a total of 10 (response rate 48%; mean 4, SD 0.9) in week 2, a total of 6 (response rate 29%; mean 4.3, SD 0.94) in week 3, and a total of 7 (response rate 33%; mean 4.4, SD 0.7) in week 4.

### Qualitative Data

Of the 21 participants, 13 (62%) participated in the 2- and 4-week interviews: 6 (46%) completed both, 4 (31%) completed only the 2-week interview, and 3 (23%) completed only the 4-week interview. Supporting quotes were identified by patient number, enrolled track (*ready to quit* or *not ready to quit*), and interview week (2 or 4 weeks).

### Engagement and Interactions With the Text Messaging Program

Participants described that they frequently read the SMS text messages:

*There were a couple of times I was in with a client or something—when I had a minute to read it, I would read it. It was never that I didn't go back to it. [P3, ready to quit, 4 weeks]*

Approximately 15% (2/13) of participants had low response rates for the SMS text message assessments. When probed for the reasons, they responded as follows:

*I responded to a couple of them and then didn't respond anymore 'cause I didn't know if it was being charged to my account. [P10, ready to quit, 4 weeks]*

*I can't read or write that much. I wait until my friend comes. When she comes, she reads them to me. Sometimes I got to wait two or three days. [P14, not ready to quit, 2 weeks]*

### Program Satisfaction

Participants were satisfied with the program, largely because they found the SMS text messages understandable:

*The thing I liked about the text messaging program is that it was straightforward. It wasn't hard for me to grasp the concept of what it was talking about. [P7, not ready to quit, 4 weeks]*

Several described how they liked being able to go back to look at the messages:

*Sometimes I would go back and look and see if there was anything helpful that could help me at the time. It was helpful. [P10, ready to quit, 4 weeks]*

Features participants thought particularly helpful included the interactive features:

*I didn't think they were going to answer that quick. For them answering quick, it helped with my craving. [P1, ready to quit, 4 weeks]*

### Perception of Program Content

#### Text Messages About Cost-Savings Associated With Stopping Smoking

Many described the cost-saving messages as helpful:

*I used to buy two packs, and now I'd buy one pack, so I'd say I'm going to smoke one, and I'd put the rest of the money in the piggy bank. Since I started with you guys, I have \$120 in my piggy bank. [P14, not ready to quit, 4 weeks]*

#### Text Messages About Provision of Resources for Unmet SDOH Needs

Some described how receiving links to resources was helpful:

*Because it gave you all the information, where to call or how to get in contact with people to try to get help...they definitely helped me out, 'cause at that time when the message came in I was low on canned food'. [P1, ready to quit, 4 weeks]*

Others preferred communicating directly with an advocate:

*I would've liked that somebody get in touch with me and to advocate to help me finish housing. [P9, ready to quit, 4 weeks]*

#### Text Messages About Managing Stress

Participants found the SMS text message tips on handling stress helpful:

*When I had the stress tips—proper breathing we do. Yeah, I found that helpful. [P1, ready to quit, 4 weeks]*

#### Text Message Framing Around the Theme: Patients Who Quit Smoking Have Higher Success in Quitting Other Substances

For some, the SMS text messages about how quitting cigarettes and opioids together could help them remain abstinent from all substances were particularly helpful:

*I like the ones for quitting other things at the same time as smoking 'cause that seems really hard for me to do. [P4, not ready to quit, 4 weeks]*

Others did not find these messages relevant as they were already on medication-assisted treatment for OUD:

*I didn't need help for that (opioids). I just needed help for smoking. [P9, ready to quit, 4 weeks]*

#### Text Messages That Provide Resources About Smoking Cessation Services

SMS text messages regarding where to find smoking cessation resources were viewed as helpful:

*It has nice little facts about smoking. You have a number that gets you on medication. And, once I tried medication, it helped me out. [P5, not ready to quit, 2 weeks]*

#### Perceived Impact of the Program on Motivation to Stop Smoking

Individuals indicated that the program was beneficial for motivating smoking cessation:

*It was like your mother in your ear, reminding you of stuff. Not in a nagging way. I was kind of surprised that it worked as well as it did but happily so. [P3, ready to quit, 4 weeks]*

Participants described that a major reason for increased motivation to stop smoking was social support:

*Sometime I was having cravings, and that moment I would receive those text message like it was telling me somebody's out there. I'm not by myself with quitting smoking. It's like I have a sponsor. [P21, ready to quit, 4 weeks]*

### **Suggestions for Improvement**

Suggestions for improvement were (1) providing supportive phone calls when needed, (2) including personal success stories, and (3) including educational videos:

*Maybe have live people to talk to when you crave something. [P5, not ready to quit, 2 weeks]*

*...to have people that have already smoked and quit have some of their personal story incorporated. [P10, ready to quit, 4 weeks]*

*I like to receive education video about quitting smoking. [P21, ready to quit, 4 weeks]*

## **Discussion**

### **Principal Findings**

Individuals with OUD meet the definition of an underserved population: they have a higher smoking prevalence than the general population, disproportionate burden of tobacco-related health disparities, increased risk factors for treatment failure, and lack of protective factors [41]. Tailored interventions for underserved populations are needed to avoid treatment failure for those seeking treatment, as well as to motivate those not ready to stop smoking. We provide evidence of the feasibility and acceptability of a newly developed SMS text messaging program for smoking cessation tailored to individuals with OUD.

Our SMS text messaging program is unique because it (1) focuses on an understudied and underserved population, (2) assesses and provides resources for unmet SDOH needs that may make the path to quitting easier, and (3) offers individualized tracks based on readiness to quit. The main findings were as follows: (1) participants reported high satisfaction with the program content and structure, and (2) participants reported that the program helped motivate smoking cessation. The vast majority of participants in our study were Medicaid insured. Medicaid recipients in Massachusetts (MassHealth members) have access to all Food and Drug Administration–approved medications, two 90-day treatment regimens per year, and 16 tobacco cessation counseling sessions per year; however, the quit rates are low. Our study adds to the growing body of evidence that SMS text messaging approaches are feasible and acceptable for providing tobacco treatment to all individuals who smoke, even among low-income Medicaid populations who have OUD and are not ready to quit.

Individuals reported that the program made them think about stopping smoking, regardless of whether they were enrolled in the *ready to quit within 30 days* or *not ready to quit within 30 days* track. SMS text messages on managing stress and providing tobacco treatment resources were perceived as particularly

helpful. Many perceived that the SMS text messages provided social support, companionship, and the motivation to stop smoking. Messages about the health benefits of quitting were well received, whereas messages on how quitting cigarettes may prevent relapse from other substances had mixed views, highlighting the importance of tailoring interventions to patient preferences.

Participants made suggestions for improvement. Some discussed how increasing the duration to 3 months would enhance the program, as would receiving supportive calls or supplemental in-person interactions as needed. Although some perceived that providing links to resources for unmet SDOH needs was adequate, others suggested that an advocate should additionally help address these needs. As suggested by the data, we plan to refine the intervention by increasing the program duration and adding supplemental in-person interactions, particularly to address unmet SDOH needs.

### **Comparison With Prior Work**

A previous intervention using SmokefreeTXT (an SMS text messaging service by the National Cancer Institute) with individuals experiencing homelessness demonstrated a median response rate of 2.1% to interactive SMS text messages, with many individuals reporting that the SMS text messages felt impersonal [42]. Our SMS text messaging program tailored to unmet SDOH needs addresses the unique circumstances of this population. Response rates to SMS text message assessments in our study ranged from 30% to 42%, with participants perceiving the program as customized to their needs.

In another study that analyzed the completion of the SmokefreeTXT program, 46% of those who set a quit date remained enrolled for the entire 42-day program. Among users who did not complete the program (eg, texted “STOP”) before program completion, the mean number of days in the program was 12 days [43]. In our study, although none of the participants dropped out of the program (eg, texted “STOP”), only 45% completed the 4-week assessments, perhaps indicating that some of these individuals did not complete the entire program. Similar findings have also been reported in other SMS text messaging interventions in underserved populations, as well as in other understudied populations [42,44,45], such as women who smoke cigarettes, where >60% of participants did not answer their phones to conduct interviews, despite multiple attempts [44].

### **Limitations**

This study has several strengths and limitations. The strengths include focusing on the understudied and underserved population of individuals with OUD and including individuals regardless of readiness to quit. These inclusions are important as evidence supports “opt-out” approaches to offering tobacco treatment to all individuals, regardless of readiness to quit [46-52]. Many studies exclude individuals with substance use disorders or psychiatric diseases [47,48,53], thus perpetuating health inequities. For qualitative studies, participants were interviewed both during and immediately after the study completion, thus minimizing recollection bias. However, our small sample size from a single recruitment site limited generalizability. Our results also reflect the findings of participants who volunteered

and thus may not reflect the perspectives of all individuals. Another limitation was that we found that a significant number of participants were ineligible during face-to-face screening as they did not have unlimited SMS text messaging or a cell phone during hospitalization. Future studies could provide cell phones or SMS text messaging plans to patients at discharge. Although our understudied population was a strength, it created a limitation for assessing feasibility and acceptability; we were unable to reach half of the participants by phone at the end of the study. This, coupled with limited resources, limited our ability to collect and measure smoking abstinence. Unlike other studies, we derived SMS text message engagement data from multiple sources (unsolicited SMS text messages, SMS text message response rates, self-report surveys, and interviews). In addition, most, if not all, SMS text message programs are limited by a lack of ability to ascertain whether the message was read. In future studies, we can additionally offer incentives to individuals responding to SMS text message assessments or

make messages more interactive through the use of quizzes or questions.

## Conclusions

SMS text messaging to promote smoking cessation may be an effective tool for improving quit rates and health outcomes in individuals who smoke cigarettes and have OUD. Our results provide valuable insights into the development and acceptability of such programs. An innovative component of our SMS text messaging intervention was screening for and providing tailored resources for unmet SDOH needs. In future studies, we will assess whether identifying unmet SDOH needs and intervening in these modifiable factors (ie, providing resources to address unmet SDOH needs) affects smoking cessation. Our next steps are to further refine the program based on patient suggestions, such as adding a community health worker or coach to address unmet SDOH needs and providing supportive phone calls when needed, and assess the effects of the refined program on smoking cessation in a randomized controlled trial.

## Acknowledgments

Financial support was provided by the Mobile and Electronic Health Affinity Research Collaborative, which is funded by the Boston University Evans Center for Interdisciplinary Biomedical Research.

## Conflicts of Interest

HK serves as a section editor for the Tobacco Dependence Treatment section for UpToDate and reports receiving personal fees from UpToDate. SW is the president of Agile Health, the company that deployed the SMS text messages described in this study.

Multimedia Appendix 1

Schedule for delivering social determinants of health assessments.

[PNG File , 328 KB - [formative\\_v6i9e36919\\_app1.png](#) ]

## References

1. Chun J, Haug NA, Guydish JR, Sorensen JL, Delucchi K. Cigarette smoking among opioid-dependent clients in a therapeutic community. *Am J Addict* 2009 Jan;18(4):316-320 [FREE Full text] [doi: [10.1080/10550490902925490](#)] [Medline: [19444736](#)]
2. Guydish J, Passalacqua E, Pagano A, Martínez C, Le T, Chun J, et al. An international systematic review of smoking prevalence in addiction treatment. *Addiction* 2016 Feb 22;111(2):220-230 [FREE Full text] [doi: [10.1111/add.13099](#)] [Medline: [26392127](#)]
3. Bandiera FC, Anteneh B, Le T, Delucchi K, Guydish J. Tobacco-related mortality among persons with mental health and substance abuse problems. *PLoS One* 2015 Mar 25;10(3):e0120581 [FREE Full text] [doi: [10.1371/journal.pone.0120581](#)] [Medline: [25807109](#)]
4. Hser Y, McCarthy WJ, Anglin M. Tobacco use as a distal predictor of mortality among long-term narcotics addicts. *Prev Med* 1994 Jan;23(1):61-69. [doi: [10.1006/pmed.1994.1009](#)] [Medline: [8016035](#)]
5. Prochaska JJ, Delucchi K, Hall SM. A meta-analysis of smoking cessation interventions with individuals in substance abuse treatment or recovery. *J Consult Clin Psychol* 2004 Dec;72(6):1144-1156. [doi: [10.1037/0022-006x.72.6.1144](#)]
6. Pakhale S, Kaur T, Charron C, Florence K, Rose T, Jama S, et al. Management and Point-of-Care for Tobacco Dependence (PROMPT): a feasibility mixed methods community-based participatory action research project in Ottawa, Canada. *BMJ Open* 2018 Jan 25;8(1):e018416 [FREE Full text] [doi: [10.1136/bmjopen-2017-018416](#)] [Medline: [29371273](#)]
7. McKelvey K, Thrul J, Ramo D. Impact of quitting smoking and smoking cessation treatment on substance use outcomes: an updated and narrative review. *Addict Behav* 2017 Feb;65:161-170 [FREE Full text] [doi: [10.1016/j.addbeh.2016.10.012](#)] [Medline: [27816663](#)]
8. Mannelli P, Wu L, Peindl KS, Gorelick DA. Smoking and opioid detoxification: behavioral changes and response to treatment. *Nicotine Tob Res* 2013 Oct 09;15(10):1705-1713 [FREE Full text] [doi: [10.1093/ntr/ntt046](#)] [Medline: [23572466](#)]
9. Kathuria H, Seibert R, Cobb V, Weinstein Z, Gowarty M, Helm E, et al. Patient and physician perspectives on treating tobacco dependence in hospitalized smokers with substance use disorders: a mixed methods study. *J Addict Med* 2019;13(5):338-345. [doi: [10.1097/ADM.0000000000000503](#)] [Medline: [31232740](#)]

10. Weinberger AH, Platt J, Esan H, Galea S, Erlich D, Goodwin RD. Cigarette smoking is associated with increased risk of substance use disorder relapse. *J Clin Psychiatry* 2017 Feb 22;78(02):e152-e160. [doi: [10.4088/jcp.15m10062](https://doi.org/10.4088/jcp.15m10062)]
11. Knudsen HK, Studts JL, Boyd S, Roman PM. Structural and cultural barriers to the adoption of smoking cessation services in addiction treatment organizations. *J Addict Dis* 2010 Jul 30;29(3):294-305 [FREE Full text] [doi: [10.1080/10550887.2010.489446](https://doi.org/10.1080/10550887.2010.489446)] [Medline: [20635279](https://pubmed.ncbi.nlm.nih.gov/20635279/)]
12. Fiore MC, Adsit R. Will hospitals finally “do the right thing”? Providing evidence-based tobacco dependence treatments to hospitalized patients who smoke. *Jt Comm J Qual Patient Saf* 2016 May;42(5):207-208. [doi: [10.1016/s1553-7250\(16\)42026-x](https://doi.org/10.1016/s1553-7250(16)42026-x)]
13. 2008 PHS Guideline Update Panel, Liaisons, Staff. Treating tobacco use and dependence: 2008 update U.S. Public health service clinical practice guideline executive summary. *Respir Care* 2008 Sep;53(9):1217-1222 [FREE Full text] [Medline: [18807274](https://pubmed.ncbi.nlm.nih.gov/18807274/)]
14. McBride C, Emmons K, Lipkus I. Understanding the potential of teachable moments: the case of smoking cessation. *Health Educ Res* 2003 Apr;18(2):156-170. [doi: [10.1093/her/18.2.156](https://doi.org/10.1093/her/18.2.156)] [Medline: [12729175](https://pubmed.ncbi.nlm.nih.gov/12729175/)]
15. Lawson PJ, Flocke SA. Teachable moments for health behavior change: a concept analysis. *Patient Educ Couns* 2009 Jul;76(1):25-30 [FREE Full text] [doi: [10.1016/j.pec.2008.11.002](https://doi.org/10.1016/j.pec.2008.11.002)] [Medline: [19110395](https://pubmed.ncbi.nlm.nih.gov/19110395/)]
16. Chu ES, Hakkarinen D, Evig C, Page S, Keniston A, Dickinson M, et al. Underutilized time for health education of hospitalized patients. *J Hosp Med* 2008 May 20;3(3):238-246. [doi: [10.1002/jhm.295](https://doi.org/10.1002/jhm.295)] [Medline: [18570334](https://pubmed.ncbi.nlm.nih.gov/18570334/)]
17. Kathuria H, Seibert RG, Cobb V, Herbst N, Weinstein ZM, Gowarty M, et al. Perceived barriers to quitting cigarettes among hospitalized smokers with substance use disorders: a mixed methods study. *Addict Behav* 2019 Aug;95:41-48. [doi: [10.1016/j.addbeh.2019.02.017](https://doi.org/10.1016/j.addbeh.2019.02.017)] [Medline: [30836208](https://pubmed.ncbi.nlm.nih.gov/30836208/)]
18. Kathuria H, Koppelman E, Borrelli B, Slatore C, Clark J, Lasser K, et al. Patient-physician discussions on lung cancer screening: a missed teachable moment to promote smoking cessation. *Nicotine Tob Res* 2020 Mar 16;22(3):431-439 [FREE Full text] [doi: [10.1093/ntr/nty254](https://doi.org/10.1093/ntr/nty254)] [Medline: [30476209](https://pubmed.ncbi.nlm.nih.gov/30476209/)]
19. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. *J Subst Abuse Treat* 2015 Sep;56:1-10. [doi: [10.1016/j.jsat.2015.01.011](https://doi.org/10.1016/j.jsat.2015.01.011)] [Medline: [25720333](https://pubmed.ncbi.nlm.nih.gov/25720333/)]
20. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2016 May 20;4(2):e49 [FREE Full text] [doi: [10.2196/mhealth.5436](https://doi.org/10.2196/mhealth.5436)] [Medline: [27207211](https://pubmed.ncbi.nlm.nih.gov/27207211/)]
21. Whittaker R, Borland R, Bullen C, Lin RB, McRobbie H, Rodgers A. Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2009 Oct 07(4):CD006611. [doi: [10.1002/14651858.CD006611.pub2](https://doi.org/10.1002/14651858.CD006611.pub2)] [Medline: [19821377](https://pubmed.ncbi.nlm.nih.gov/19821377/)]
22. Borrelli B, Bartlett YK, Tooley E, Armitage CJ, Wearden A. Prevalence and frequency of mHealth and eHealth use among US and UK smokers and differences by motivation to quit. *J Med Internet Res* 2015 Jul 04;17(7):e164 [FREE Full text] [doi: [10.2196/jmir.4420](https://doi.org/10.2196/jmir.4420)] [Medline: [26149323](https://pubmed.ncbi.nlm.nih.gov/26149323/)]
23. Mobile fact sheet. Pew Research Center. 2021 Apr 7. URL: <https://www.pewresearch.org/internet/fact-sheet/mobile/> [accessed 2022-02-19]
24. Chilukuri N, West M, Henderson JL, Lawson S, Ehsanipoor R, Costigan K, et al. Information and communication technology use among low-income pregnant and postpartum women by race and ethnicity: a cross-sectional study. *J Med Internet Res* 2015 Jul 03;17(7):e163 [FREE Full text] [doi: [10.2196/jmir.3916](https://doi.org/10.2196/jmir.3916)] [Medline: [26142162](https://pubmed.ncbi.nlm.nih.gov/26142162/)]
25. Abrams L, Wu K, Krishnan N, Long M, Belay S, Sherman S, et al. A pilot randomized controlled trial of text messaging to increase tobacco treatment reach in the emergency department. *Nicotine Tob Res* 2021 Aug 18;23(9):1597-1601. [doi: [10.1093/ntr/ntab036](https://doi.org/10.1093/ntr/ntab036)] [Medline: [33684207](https://pubmed.ncbi.nlm.nih.gov/33684207/)]
26. Marmot M. Social determinants of health inequalities. *Lancet* 2005 Mar;365(9464):1099-1104. [doi: [10.1016/s0140-6736\(05\)71146-6](https://doi.org/10.1016/s0140-6736(05)71146-6)]
27. WHO Healthy Cities Project, WHO International Centre for Health and Society. Social Determinants of Health The Solid Facts. Europe: World Health Organization, Regional Office for Europe; 2003.
28. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep* 2014 Jan 01;129(1\_suppl2):19-31. [doi: [10.1177/00333549141291s206](https://doi.org/10.1177/00333549141291s206)]
29. Sheira LA, Frongillo EA, Hahn J, Palar K, Riley ED, Wilson TE, et al. Relationship between food insecurity and smoking status among women living with and at risk for HIV in the USA: a cohort study. *BMJ Open* 2021 Sep 06;11(9):e054903 [FREE Full text] [doi: [10.1136/bmjopen-2021-054903](https://doi.org/10.1136/bmjopen-2021-054903)] [Medline: [34489299](https://pubmed.ncbi.nlm.nih.gov/34489299/)]
30. Kim-Mozeleski JE, Seligman HK, Yen IH, Shaw SJ, Buchanan DR, Tsoh JY. Changes in food insecurity and smoking status over time: analysis of the 2003 and 2015 panel study of income dynamics. *Am J Health Promot* 2019 Jun 21;33(5):698-707 [FREE Full text] [doi: [10.1177/0890117118814397](https://doi.org/10.1177/0890117118814397)] [Medline: [30463414](https://pubmed.ncbi.nlm.nih.gov/30463414/)]
31. Armour BS, Pitts MM, Lee C. Cigarette smoking and food insecurity among low-income families in the United States, 2001. *Am J Health Promot* 2008 Jul 01;22(6):386-392. [doi: [10.4278/ajhp.22.6.386](https://doi.org/10.4278/ajhp.22.6.386)] [Medline: [18677878](https://pubmed.ncbi.nlm.nih.gov/18677878/)]

32. HUD 2019 continuum of care homeless assistance programs homeless populations and subpopulations. U.S. Department of Housing and Urban Development. 2019. URL: [https://files.hudexchange.info/reports/published/CoC\\_PopSub\\_NatlTerrDC\\_2019.pdf](https://files.hudexchange.info/reports/published/CoC_PopSub_NatlTerrDC_2019.pdf) [accessed 2022-05-10]
33. Blanco C, Wiley TR, Lloyd JJ, Lopez MF, Volkow ND. America's opioid crisis: the need for an integrated public health approach. *Transl Psychiatry* 2020 May 28;10(1):167 [FREE Full text] [doi: [10.1038/s41398-020-0847-1](https://doi.org/10.1038/s41398-020-0847-1)] [Medline: [32522999](https://pubmed.ncbi.nlm.nih.gov/32522999/)]
34. Herbst N, Wiener RS, Helm ED, O'Donnell C, Fitzgerald C, Wong C, et al. Effectiveness of an opt-out electronic health record-based tobacco treatment consult service at an urban safety net hospital. *Chest* 2020 Oct;158(4):1734-1741. [doi: [10.1016/j.chest.2020.04.062](https://doi.org/10.1016/j.chest.2020.04.062)] [Medline: [32428510](https://pubmed.ncbi.nlm.nih.gov/32428510/)]
35. Smokefree Text Messaging Programs. [smokefree.gov](https://smokefree.gov/tools-tips/text-programs). URL: <https://smokefree.gov/tools-tips/text-programs> [accessed 2022-02-19]
36. Borrelli B, Henshaw M, Endrighi R, Adams WG, Heeren T, Rosen RK, et al. An interactive parent-targeted text messaging intervention to improve oral health in children attending urban pediatric clinics: feasibility randomized controlled trial. *JMIR Mhealth Uhealth* 2019 Nov 11;7(11):e14247 [FREE Full text] [doi: [10.2196/14247](https://doi.org/10.2196/14247)] [Medline: [31710306](https://pubmed.ncbi.nlm.nih.gov/31710306/)]
37. Buitron de la Vega P, Losi S, Sprague Martinez L, Bovell-Ammon A, Garg A, James T, et al. Implementing an EHR-based screening and referral system to address social determinants of health in primary care. *Med Care* 2019 Jun;57 Suppl 6 Suppl 2:S133-S139. [doi: [10.1097/MLR.0000000000001029](https://doi.org/10.1097/MLR.0000000000001029)] [Medline: [31095052](https://pubmed.ncbi.nlm.nih.gov/31095052/)]
38. Heatherton T, Kozlowski L, Frecker R, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. *Br J Addict* 1991 Sep;86(9):1119-1127. [doi: [10.1111/j.1360-0443.1991.tb01879.x](https://doi.org/10.1111/j.1360-0443.1991.tb01879.x)] [Medline: [1932883](https://pubmed.ncbi.nlm.nih.gov/1932883/)]
39. Fagerström K. Determinants of tobacco use and renaming the FTND to the Fagerstrom test for cigarette dependence. *Nicotine Tob Res* 2012 Jan 24;14(1):75-78. [doi: [10.1093/ntr/ntr137](https://doi.org/10.1093/ntr/ntr137)] [Medline: [22025545](https://pubmed.ncbi.nlm.nih.gov/22025545/)]
40. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015 Mar 11;3(1):e27 [FREE Full text] [doi: [10.2196/mhealth.3422](https://doi.org/10.2196/mhealth.3422)] [Medline: [25760773](https://pubmed.ncbi.nlm.nih.gov/25760773/)]
41. Borrelli B. Smoking cessation: next steps for special populations research and innovative treatments. *J Consult Clin Psychol* 2010 Feb;78(1):1-12. [doi: [10.1037/a0018327](https://doi.org/10.1037/a0018327)] [Medline: [20099945](https://pubmed.ncbi.nlm.nih.gov/20099945/)]
42. Baggett TP, McGlave C, Kruse GR, Yaqubi A, Chang Y, Rigotti NA. SmokefreeTXT for homeless smokers: pilot randomized controlled trial. *JMIR Mhealth Uhealth* 2019 Jun 04;7(6):e13162 [FREE Full text] [doi: [10.2196/13162](https://doi.org/10.2196/13162)] [Medline: [31165717](https://pubmed.ncbi.nlm.nih.gov/31165717/)]
43. Prutzman YM, Wiseman KP, Grady MA, Budenz A, Grenen EG, Vercammen LK, et al. Using digital technologies to reach tobacco users who want to quit: evidence from the national cancer institute's smokefree.Gov initiative. *Am J Prev Med* 2021 Mar;60(3 Suppl 2):S172-S184 [FREE Full text] [doi: [10.1016/j.amepre.2020.08.008](https://doi.org/10.1016/j.amepre.2020.08.008)] [Medline: [33663705](https://pubmed.ncbi.nlm.nih.gov/33663705/)]
44. Sloan M, Hopewell S, Coleman T, Cooper S, Naughton F. Smoking cessation support by text message during pregnancy: a qualitative study of views and experiences of the MiQuit intervention. *Nicotine Tob Res* 2017 May 01;19(5):572-577 [FREE Full text] [doi: [10.1093/ntr/ntw241](https://doi.org/10.1093/ntr/ntw241)] [Medline: [28403457](https://pubmed.ncbi.nlm.nih.gov/28403457/)]
45. Alcaraz KI, Riehman K, Vereen R, Bontemps-Jones J, Westmaas JL. To text or not to text? Technology-based cessation communication preferences among urban, socioeconomically disadvantaged smokers. *Ethn Dis* 2018 Jul 12;28(3):161-168 [FREE Full text] [doi: [10.18865/ed.28.3.161](https://doi.org/10.18865/ed.28.3.161)] [Medline: [30038477](https://pubmed.ncbi.nlm.nih.gov/30038477/)]
46. Faseru B, Ellerbeck EF, Catley D, Gajewski BJ, Scheuermann TS, Shireman TI, et al. Changing the default for tobacco-cessation treatment in an inpatient setting: study protocol of a randomized controlled trial. *Trials* 2017 Aug 14;18(1):379 [FREE Full text] [doi: [10.1186/s13063-017-2119-9](https://doi.org/10.1186/s13063-017-2119-9)] [Medline: [28806908](https://pubmed.ncbi.nlm.nih.gov/28806908/)]
47. Fu SS, van Ryn M, Sherman SE, Burgess DJ, Noorbaloochi S, Clothier B, et al. Proactive tobacco treatment and population-level cessation: a pragmatic randomized clinical trial. *JAMA Intern Med* 2014 May 01;174(5):671-677. [doi: [10.1001/jamainternmed.2014.177](https://doi.org/10.1001/jamainternmed.2014.177)] [Medline: [24615217](https://pubmed.ncbi.nlm.nih.gov/24615217/)]
48. Nahhas GJ, Wilson D, Talbot V, Cartmell KB, Warren GW, Toll BA, et al. Feasibility of implementing a hospital-based "Opt-out" tobacco-cessation service. *Nicotine Tob Res* 2017 Aug 01;19(8):937-943. [doi: [10.1093/ntr/ntw312](https://doi.org/10.1093/ntr/ntw312)] [Medline: [27928052](https://pubmed.ncbi.nlm.nih.gov/27928052/)]
49. Ylioja T, Reddy V, Ambrosino R, Davis EM, Douaihy A, Slovenkay K, et al. Using bioinformatics to treat hospitalized smokers: successes and challenges of a tobacco treatment service. *Jt Comm J Qual Patient Saf* 2017 Dec;43(12):621-632. [doi: [10.1016/j.jcjq.2017.06.010](https://doi.org/10.1016/j.jcjq.2017.06.010)] [Medline: [29173282](https://pubmed.ncbi.nlm.nih.gov/29173282/)]
50. Haas JS, Linder JA, Park ER, Gonzalez I, Rigotti NA, Klinger EV, et al. Proactive tobacco cessation outreach to smokers of low socioeconomic status: a randomized clinical trial. *JAMA Intern Med* 2015 Feb 01;175(2):218-226 [FREE Full text] [doi: [10.1001/jamainternmed.2014.6674](https://doi.org/10.1001/jamainternmed.2014.6674)] [Medline: [25506771](https://pubmed.ncbi.nlm.nih.gov/25506771/)]
51. Cartmell K, Dismuke C, Dooley M, Mueller M, Nahhas G, Warren G, et al. Effect of an evidence-based inpatient tobacco dependence treatment service on 1-year postdischarge health care costs. *Med Care* 2018 Oct;56(10):883-889 [FREE Full text] [doi: [10.1097/MLR.0000000000000979](https://doi.org/10.1097/MLR.0000000000000979)] [Medline: [30130271](https://pubmed.ncbi.nlm.nih.gov/30130271/)]
52. Cartmell K, Dooley M, Mueller M, Nahhas G, Dismuke C, Warren G, et al. Effect of an evidence-based inpatient tobacco dependence treatment service on 30-, 90-, and 180-day hospital readmission rates. *Med Care* 2018 Apr;56(4):358-363 [FREE Full text] [doi: [10.1097/MLR.0000000000000884](https://doi.org/10.1097/MLR.0000000000000884)] [Medline: [29401186](https://pubmed.ncbi.nlm.nih.gov/29401186/)]

53. Vick L, Duffy S, Ewing L, Rugen K, Zak C. Implementation of an inpatient smoking cessation programme in a Veterans Affairs facility. *J Clin Nurs* 2013 Mar;22(5-6):866-880. [doi: [10.1111/j.1365-2702.2012.04188.x](https://doi.org/10.1111/j.1365-2702.2012.04188.x)] [Medline: [22882776](https://pubmed.ncbi.nlm.nih.gov/22882776/)]

## Abbreviations

**BMC:** Boston Medical Center

**OD:** opioid use disorder

**SDOH:** social determinants of health

**THRIVE:** Tool for Health and Resilience In Vulnerable Environments

*Edited by A Mavragani; submitted 31.01.22; peer-reviewed by L Popova, C Memering; comments to author 27.06.22; revised version received 17.07.22; accepted 28.07.22; published 01.09.22.*

*Please cite as:*

*Kathuria H, Shankar D, Cobb V, Newman J, Bulekova K, Wertz S, Borrelli B*

*Integrating Social Determinants of Health With Tobacco Treatment for Individuals With Opioid Use Disorder: Feasibility and Acceptability Study of Delivery Through Text Messaging*

*JMIR Form Res* 2022;6(9):e36919

URL: <https://formative.jmir.org/2022/9/e36919>

doi: [10.2196/36919](https://doi.org/10.2196/36919)

PMID: [36048509](https://pubmed.ncbi.nlm.nih.gov/36048509/)

©Hasmeena Kathuria, Divya Shankar, Vinson Cobb, Julia Newman, Katia Bulekova, Scott Wertz, Belinda Borrelli. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 01.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# The Impact of Smartphone Apps Designed to Reduce Food Waste on Improving Healthy Eating, Financial Expenses and Personal Food Waste: Crossover Pilot Intervention Trial Studying Students' User Experiences

Therese Fostervold Mathisen<sup>1</sup>, BSc, MSc, PhD; Frode Ramstad Johansen<sup>2</sup>, BSc, MSc, PhD

<sup>1</sup>Faculty of Health, Welfare and Organisation, Østfold University College, Halden, Norway

<sup>2</sup>Faculty of Computer Sciences, Engineering and Economics, Østfold University College, Halden, Norway

**Corresponding Author:**

Therese Fostervold Mathisen, BSc, MSc, PhD

Faculty of Health, Welfare and Organisation

Østfold University College

Høgskolen i Østfold

PO Box 700

Halden, No-1757

Norway

Phone: 47 69 60 80 00

Fax: 47 69 60 80 01

Email: [therese.f.mathisen@hiof.no](mailto:therese.f.mathisen@hiof.no)

## Abstract

**Background:** Global sustainability and individual health need coordinated attention. While individuals are recommended a healthy diet to reduce the burden of noncommunicable diseases, global attention to natural resource conservation is also needed. The latter specifically means effective measures to reduce food waste.

**Objective:** This pilot study evaluates the experiences of students and effect from using smartphone apps designed to reduce food waste on personal healthy eating, financial expenses, and food waste.

**Methods:** A total of 6 students from different study programs (mean age 24.7, SD 2.9) were recruited to evaluate 2 different apps designed to reduce food waste and to register food consumption, food waste, and financial food expenses before and after the app trials. The apps evaluated were the commercially available TotalCtrl Home and Too-Good-To-Go. Results were analyzed by mixed methods, comprising statistical analyses for quantifiable data and thematic analyses for qualitative data. The apps were used separately in random order, each for 1 month. Primary outcome was user expectations to and experiences from the use of the apps, which were obtained by semistructured interviews. Secondary outcomes were changes in food waste volume, financial food expenses, and healthy eating. While information on food waste and food expenses was obtained by weighing food waste and registering food costs for 2 weeks before and after app trials, scores for consuming healthy diets were calculated from registered food records by scoring criteria matched to national recommendations for healthy eating.

**Results:** Awareness on food waste increased after app trials, but experiences with apps pointed toward several potential for technical and content improvements. The students reported that there were too many manual operations in the apps to induce permanent use (TotalCtrl Home), that services seemed more concerned about the producers' interests than the individual's needs (Too-Good-To-Go), and that they missed a composite app that included functions to promote healthy eating and overview of budget and expenses as well as of food waste (both apps). Use of apps designed to reduce food waste and personal costs and to improve healthy eating did not result in any measurable effects, that is, no change in food waste (mean change 0.81, SD 1.5 kg;  $P=.13$ ), healthy eating (mean change  $-0.24$ , SD 0.43;  $P=.24$ ), or personal food expenses (mean change 47.5 NOK or US \$4.8, SD 416.9 NOK or US \$42.5;  $P=.39$ ).

**Conclusions:** Apps may aid in increased awareness of food waste at the producer and consumer levels. Large-scale studies with longer duration are needed to see if apps may induce measurable changes in food waste, healthy eating, and financial expenses.

(JMIR Form Res 2022;6(9):e38520) doi:[10.2196/38520](https://doi.org/10.2196/38520)



**KEYWORDS**

smartphone app; food waste; healthy eating; diet; automatic; registration; global sustainability; financial expenses

## *Introduction*

Food waste means the loss of produced food meant for consumption by humans, and is estimated to encompass about 20%-30% of all produced food [1-3]. The loss of edible food means loss of important resources, such as energy and capital, which puts a constant stress on natural resources (involves all stages of the product life cycle, from production to destruction), including the production of greenhouse gases that cause global warming [4-7]. To counteract these severe consequences, effective measures to reduce food waste are requested, and the European Parliament has put forth a goal to reduce food waste by 50% within 2030 [7].

A significant amount of food waste occurs at the individual consumer level, with an estimated loss of 179 kg per capita per year in the European Union [8-10]. As such, motivating the individual for making personal efforts in reducing food waste is essential. Importantly, study findings suggest that individuals are prone to commit to food waste-reducing behavior if they are educated about the consequences (acting to reduce feeling of guilt), and if they feel expectation and joint commitment from the society [11]. On the contrary, if they have a busy time schedule and experience the food waste-reduction behavior to require extra efforts, they are less likely to engage [11,12]. Younger generations may be less knowledgeable about reuse of leftover foods, and as such engage in higher food waste behavior than older generations who are typically more educated on food management from their experience of periods with shortness of food supply or limited personal finances [13,14]. Nevertheless, young adults are highly aware of the environmental consequences from our consumer culture in general, and as such may be motivated to change behavior if making an impact. If efforts concurrently favor their own economic situation, organized systems easing the efforts to reduce food waste could enhance their motivation to engage [15].

Today, young adults are massive users of digital technology, with over 95% in the advanced economies being smartphone users [16]. Smartphone apps have become convenient programs for many daily activities, and are perfect channels to deliver personalized and socially responsible interventions [17]. Previous research on apps for health or behavioral interventions among young adults reported that the participants did not find the apps personally relevant, but revealed tentative willingness to test them [18]. Hence, it is important that the apps have features that appeal to the young consumer. A recent structured review of available apps to promote sustainable waste management behaviors identified the 6 most persuasive

strategies applied with the aim of influencing the users' behavior [17]. The first prominent features listed were a reduction of complex app tasks to make the use of the app easier and a personalized content. Next, an experience of a real-world belongingness by bringing information about the app company and surface credibility (ie, competent look and use) were listed as typical. Lastly, functions such as reminders and self-monitoring were provided in an attempt to motivate and engage the user. Self-monitoring may be specifically reinforcing when payoff is measurable and real, and not only considerations for the environment, but also personal profits could embody such achievements. For young adults such as students, financial restraints often limit their leisure spending and participation. If education and awareness about the impact of slight modification of everyday behavior can make them realize financial savings, they may improve motivation to engage in such a behavior.

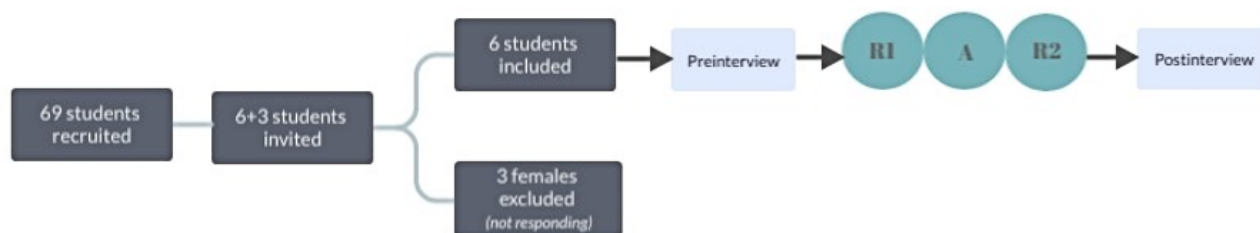
Importantly, the elements typically employed in apps for behavioral changes relate to the 3 factors highlighted as necessary to assist behavioral change: motivation, ability, and triggers [17,19]. As such, apps may emerge as promising aids for the individual to comply with the intention to reduce food waste and concurrently experience personal rewards such as saving money. With such intentions in mind, the health of the individual must also be taken into consideration. Individuals in the modern society need to improve their diets to expect healthy life [20,21]; hence, the consideration for the environment (ie, reduce food waste) and motivation for financial savings must include attention to the personal diet. The intention of this pilot study was to explore whether smartphone apps can help students to reduce their personal food waste, improve their habitual intake of a healthy and varied diet, and reduce financial expenses on food.

## *Methods*

### **Design and Procedure**

This is a pilot intervention study exploring students' expectations to and experiences from using smartphone apps that aim at reducing food waste and evaluating its effects on food waste, healthy eating, and financial food expenses. As such, this study combines efficacy and usability testing, and provides a broader perspective and evaluation as previously highlighted in a viewpoint paper [22]. The study included a preperiod with baseline registrations and interviews on expectations, a trial period in which 2 different apps were explored for 1 month each by all the participants in a random crossover order, and a final postperiod repeating the baseline registrations and interviews on experiences (Figure 1).

**Figure 1.** Study flow. Three females did not respond to invitation, and as such three more females from the recruited pool were invited. In total 6 students (3 males and 3 females) consented to participate. Note: Preinterview, individual interviews on expectations. R1, 2-week preregistration period on food waste, diet and dietary expenses. A, two-month period trying two digital phone applications consecutively. R2, 2-week postregistration period. Postinterview, individual interviews on experiences.



## Participants

We recruited students from a university in south east of Norway during the winter of 2020/2021. Recruitment messages were distributed through the university's internal communication system (Canvas), asking for students interested in digital interventions to reduce food waste, and rewarding the 3 months' participation with a financial compensation of 7500 NOK (US \$765). In total, 69 students responded, from which a total of 9 were invited, and 6 were finally included (Figure 1). Students were sorted by sex, housing condition, study, and experiences with food waste reduction before selected invitations were sent (from the pool of recruited participants) to present a group with balanced characteristics.

## Outcomes

This pilot study followed a mixed methods design, including quantifiable and qualitative outcome data.

All participants were instructed to manually register their food consumption, food expenses, and financial expenses for food in the 2 weeks before and after the app trial period. The registration was by use of a precoded MS Excel (Microsoft, Inc.) document. The information and instructions on registrations and use of registration tools (ie, kitchen scale and excel document) were provided in a synchronous digital meeting, and any expenses related to the procurement of the kitchen scale were covered by the project sponsor. Food consumption was registered by noting details on time for consumption, type and volume of food consumed, and finally noting any leftovers not consumed (added to the total registration of foods wasted from the private household). Foods wasted were noted separately in the MS Excel document after weighing foods before they were thrown in the litter. All expenses for food, including receipts from shopping, were registered in the same MS Excel document on a separate sheet. Foods consumed were evaluated according to a healthy index, foods wasted were evaluated according to weighed volume, and food expenses were compared before and after the app trial solely by calculation comparisons.

The healthy food index created for this study was kept concealed from participants to avoid biases, such as changes occurring simply from the awareness of being evaluated. The index consisted of a total score ranging from 0 to 5, and was calculated on the following premises: 1 point if more than half of the daily grain/grain products consumed were whole grain products; 0.5 points for each portion of the "5-a-day" recommendation for

fruits and vegetables (each portion being 100 g, and providing a maximum of 2.5 points); 0.5 points if sugary drinks were absent/nonsugary drinks were preferred; and 1 point for consuming fish. These criteria match the international and national recommendations for a healthy diet, highlighting the health benefits of increasing the intake of plant-based foods and fiber, reducing intake of sugar, and by replacing fish for meat [20,23].

Participants were interviewed following a semistructured manual once before the trial, focusing on previous experiences on app use to reduce food waste, increase healthy eating, or control economy, and on expectations of the intervention period. Finally, the participants were interviewed after the intervention about their experiences from using apps to reduce food waste and increase healthy eating, as well as to understand how these affected their financial expenses. See [Multimedia Appendix 1](#) for interview guides. Both authors studied the transcribed interviews (in Norwegian) and approved the English translation of quotes. Translation of quotes was done with the intention to reflect the oral spoken language by the participants (ie, not primarily considering the optimal grammatical form of expression/written language).

## Intervention: Smartphone Apps

The 2 apps used in this trial were Too-Good-To-Go (TGTG) and TotalCtrl Home. Each app was designed to aid in reducing food waste and to provide the user with a chance for reduced food expenses.

TGTG covers major European cities and is an app designed to help restaurants and stores with unsold food surplus to reach potential customers. As such, the stores may be able to reduce food waste and also have a small income from the surplus sales. The stores may post their food surplus in the app with a reduced price and with a defined time frame in which customers may bid for and collect the items. The products are reserved by the customer when TGTG confirms their order with a reservation confirmation. The customer has the possibility to filter the results in the app in consideration of the availability of products to reserve, of the pickup hour, or of the nature of products contained in "magic bags." The pickup time will normally be in a period of 10-30 minutes but can be both shorter and longer. The profit for the customer is obvious, a potential to reduce personal food spending by having access to price-reduced foods and by aiding in reducing global food waste.

TotalCtrl Home is an app designed to aid in food waste reduction at the individual level, to inspire cooking and as such to facilitate better financial utilization of food investments. The functions are about providing an overview of the food content in the home, to offer recipes based on the foods registered in the personal app “kitchen” and specifically for those food items registered with an expiry date, and to assist with the creation of a shopping list based on suggested menus included in the app register. New food items procured for the household may be registered in the app by scanning the barcode or by manually typing in the details, resulting in an addition to the app kitchen. The foods may be registered as stored in the freezer, the refrigerator, or the cabinets, and must be manually updated if changes occur (ie, used for cooking, changing storing location, wasted).

### Statistical Analysis

The quantitative data were analyzed using SPSS Statistics version 27 (IBM, Inc.). The healthy food index (total score), amount of food wasted (in grams), and personal financial expenses were analyzed for any changes from pre- to posttrial. With regard to the small group of participants, and the robustness of the Student *t* test, such parametric analyses were applied to explore the pre-post changes.

The interviews were analyzed thematically as suggested by Braun and Clarke [24]: transcribing and familiarizing with the data, generating initial codes for the data, identifying and reviewing themes by collated codes, naming themes, and producing the report. During the final stage, rich extracts were chosen for illustration of themes, and finally, the analyses were checked according to the original research question.

### Ethics Approval

This trial and its data collection and storage plan have been approved by the Norwegian Agency for Shared Services in Education and Research on April 3, 2021 (id number 832471). All included participants signed an informed consent before participation in this trial.

## Results

### Overview

The group of 6 recruited students had a mean age of 24.7 (SD 2.9) years and represented 6 different study programs (nursing, information technology, teaching, digital media and design, data science, and economy and administration), and were equally distributed with regard to gender. All participants lived in a single-person housing unit, and reported to exercise a mean number of 2.7 (SD 1.5) sessions per week. As many as 5 of the 6 participants reported to have previously tried some initiatives to reduce food waste, and 5 participants reported to have tried apps to register and analyze their diet.

### Part 1: Expectations About Participating in a Project Exploring the Effect of Apps Aiming to Reduce Food

### Waste, Contribute to a Healthier Diet, and Save Money on Foods

#### *Personal Needs Before Global Needs*

The students were first asked about their motivation for attending this project. Most students indicated the need to reduce personal expenses for food as the most important motive to engage in reduction of food waste. One student talked about her routine of always turning to the desk with last-minute offers in the grocery store, because there often are affordable offers. Having her friend doing the same thing kind of normalized that behavior to her. The promise of saving money by using an app and the recommendations from relatives and friends about all the cheap food made available through an app were important motives reported by most students. But few had continued to use the apps, because of less positive experiences.

*It's mainly about the economy. How much can you save by reducing personal food waste, or... Well, for some, saving money is more motivating than reducing food waste due to environmental issues. [Male student, 25 years]*

*...you get a lot of bakery goods, and I know I'm not going to be eating all those wheat buns. And then... I don't know, well, they end up in the freezer and then you eat them some other time you'd fancy some bakery goods. But this is the main reason I do not use those apps any more. They sort of..., well, yes, you do save a lot of money on those eight wheat buns, but you weren't going to buy eight wheat buns in the first place. [Female student, 22 years]*

Other than the experiences of being offered bakery goods for snacks rather than foods for main meals, the short time slots and competition to be in reach of the offers were experienced as a large limitation to engage further. The experiences from being primarily offered snacks further draw attention to the personal health perspective when joining such app initiatives. One student told about her initial positive experiences on using apps to reduce food waste, as this made her save a lot of money. But as her awareness on personal health and reasonable body weight regulation increased, she started looking for other apps.

*Everything starts from myself, like what I want first. I just want... I don't like my body to feel out of control, so that's why I try to control myself, and to be controlling it, I need some tools. [Female student, 27 years]*

This female student was very aware of the global issue with food shortages, and was raised with a view of food as something to praise. The student really wanted to explore new foods, but yet determined not to produce unnecessary waste caused by personal pickiness. One way to accommodate these contradictions was to use an app where food stores could offer cheaper last-minute foods. The student first found this as a nice way to try new foods, while concurrently helping the stores reduce food waste and have income for these last-minute offers. However, as the student increased her awareness of healthy foods, the first positive experiences vanished with several

disappointments (ie, foods being trashed because they were not suitable for the student's needs).

### **Global Awareness Challenges Personal Pleasure**

When the students were asked what they thought could be a meaningful personal contribution to reduce food waste, they all responded with suggestions related to personal cooking and shopping behavior. They mentioned that they made sure leftovers from cooking and meals were reused and not thrown away in the litter. They argued to plan the meals, to know about and wisely use what is present in the refrigerator, and to save any leftover from meals for later use. Attention to what is bought for the household, by knowing what is already in the refrigerator and what is actually needed, was another suggestion made by most students. One student reflected over her awareness of food waste from the personal household:

*Well, I live alone, which means that all I put into my refrigerator must be consumed by me...or at least I must in some practical way make sure it is consumed. And then you have these fantastic, green recyclable bin-bags; where you see exactly the volume of food you trash. [Female student, 22 years]*

One student also argued that moderation would be an important move to reduce food waste in a global perspective—simply not to overconsume food. The overconsumption by single individuals does not only challenge personal health, but also the global availability of natural resources. Implicit in the reflections by these students is also a premise of purchasing only what you as a single individual or personal household would need and what you know you would be able to consume. Referring to the latter, one student reflected on this, finding the global consideration to be a personal limitation:

*I also want to try a new food, a new type of food; just for the experience. And maybe it just might taste better. But it may sometimes cost me more. Because, you know, if I try a new pie, and then if I don't like it, then probably I'm going...not going to eat it. [Female student, 27 years]*

### **Artificial Intelligence, and Not Too Detailed, Please**

All but 2 had already tried different types of apps to reduce food waste, or apps to increase insight into personal diet and health. While the former was primarily motivated by potential personal economic savings, the latter was by different reasons. Still, while the health motives varied from sport performance, body weight control, to illness control, they all agreed that detailed nutritional information was not in their interest, or could induce issues with food:

*No, I mean...I think it brings too much stress. I cannot bear to be that detailed. Really, I want to live a normal, balanced life. [Male student, 25 years]*

One student had also wanted to find an app to gain control over her personal financial expenses. Here too, details, advanced analyses and functions were not needed, specifically if these additions meant charges for use. As such, she chose to use a simple MS Excel program to log her incomes and expenses. Her system, including logs of economy and logs of nutritional

intake, with the aim to keep low food waste, resulted in the use of multiple systems that needed to be operated manually.

*That's when I searched for the applications that could help me to gain control, and control is not only about health or finance, but also my lifestyle and my time. Because if I do something manually, it will be really time consuming. Like, it's automatic. I don't have to type in things; that save time and is more convenient. [Female student, 27 years]*

The idea that apps should be intelligent and demand less work and manual operations was mentioned by most students. Further, an app should not be too advanced with too many functions or information, but rather provide an easy overview of financial expenses, energy intake, or food waste. Such overviews could be comparisons on a weekly or monthly basis, to see if one keeps track of low food waste or financial savings. Both the perspective of time use and the burden experienced by manually operating the apps were mentioned. Just as important was the preparatory work one has to do to provide information to the app, such as if one has to weigh the food consumed for dietary calculations, or the food *not* consumed for food waste calculations. Several mentioned the idea of taking pictures of items, or scanning a barcode, and that this should provide the app with enough information to do the necessary registration and calculations. Additionally, an app that could aid as an extended source of memory and creativity was held as desirable. It should remember what you already have at home when you are in the grocery store, should suggest menus based on your food stock, and should automatically calculate the nutritional content in what you cook and consume.

## **Part 2: Experiences From Using Apps to Reduce Food Waste, Improve Diet, and Gain Financial Control**

### **Main Experience**

After a period of 2 months, none of the students felt the apps had helped them in reducing their food waste noticeably. However, most felt their *awareness* of food waste had improved, and that there had been an economic impact from better planning (eg, knowing about your food stock before going to the grocery store, avoiding impulsive food shopping, and by reusing leftover foods).

### **Big Brother**

The awareness of food waste, both at the individual level and at the wholesaler and producer levels, was reported to be brought into everyday attention simply by joining this project. But the functions provided by the apps were also reported as effective tools. The experience from viewing repeated offers in the TGTG app increased the understanding of the volume of food wasted at the producer and wholesaler levels. Also, the registration of foods in the TotalCtrl app augmented the attention of food in personal stock.

*I feel that I have a larger responsibility in reducing food waste after being made aware of how much food is wasted from the different wholesalers. I could easily be waiting a few more minutes at the store after picking up my pre-booked bag if I notice there are*

*more bags to be saved from being wasted. ( - ) It's actually ok to realize that I can have this responsibility; I mean, if its only about me eating crisp bread rather than my usual bread, I'm more than willing to do that. [Female student, 22 years, on experiences from TGTG]*

*This project has made me more aware of due dates. I look for due dates maybe once per week in my fridge; are there any items I need to plan to use, in order to consume them before the due dates. [Female student, 23 years, on experiences from TotalCtrl]*

The students further mentioned the pre- and postregistration period to be eye-openers not only to personal food waste, but also to the economic potential and personal health.

*It could be incidental, but it is likely that when you register your diet, and the registrations say pizza and burgers for four days in a row... (-) You become more aware of it, when you see it like that, written in words. You end up thinking; this is not good for you. [Male student, 29 years, on experiences from pre- and postfood registration]*

Nevertheless, the scenario of being recruited to a study and feeling monitored may cause a Hawthorne effect, in which the participants are very aware of their behavior and of the intention of the study. As such, it may be difficult to say if the efforts made (eg, reducing food waste) are an effect from app use or from being observed and monitored.

*It's a bit like...when you know you are being observed, then you want to perform, and as such I think I have consumed healthier food. (-) But I think the app is there for the better. It will be kept in my phone. I definitely will...continue to re-use food leftovers in other meals. I have learned new things during this period. [Female student, 23 years, on experiences from TotalCtrl]*

*And as far as I understood that during the period we used the app, I actually had much better control over the fridge and everything that was there. [Male student, 22 years, on experiences from TotalCtrl]*

Considering that most students wanted to continue using the apps after the study, it seems logical to say that the apps are helpful tools. This means that consumers first need to be recruited for app use to realize the potential within them, and then to find routines to continue using them during hectic periods rather than to fall back to old behaviors.

*It's mostly about personal motivation. That...kind of, the busier you are, the more likely such new things fade out. If you have short time and need to cook dinner quickly, then you'll just reach for something simple on your way home. You just forget about it in a hectic everyday life. [Male student, 22 years]*

### **Food Waste: The Consumer or The Producer Perspective**

One of the main objectives reported by the students about joining a study aiming to reduce food waste was to experience economic benefits. While a student said she had noticed that

her receipts were getting shorter by using the apps, few said they had experienced any favorable economic effect. Most reported first to be fascinated by the cheap offers made with the TGTG, but to experience increased personal food waste and uneconomical spending in the long run. To be effective at reducing food waste, students argued the apps had to facilitate choices from several wholesalers in the living area and not only kiosks and bakeries, to not be confined to limited pickup times, and to possess a personal willingness to eat what is offered rather than according to a planned diet or personal likings.

*I'm very picky when it comes to food that is due date or approaching expiration date, and a bit picky with food in general, and if you never know what you're getting in a bag like that then...much is thrown in the litter anyway. [Male student, 29 years, on experiences from TGTG]*

*Sometimes I cannot consume the foods because of my diet. So, I have to give it away. But if I cannot find someone to give it away to, I have to throw it away. So, the apps sometimes also increased my food waste. [Female student, 27 years, on experiences from TGTG]*

*Its definitely cheap, but I feel like I'm...wasting money when I do not want to eat what I get. [Male student, 25 years, on experiences from TGTG]*

Most students argued that the TGTG moved the issue with food waste from the wholesale to the consumer level, as the offers were mostly about bakery goods in larger amounts than a person would repeatedly be willing to buy and eat. As such, most had discontinued the use of TGTG other than on occasions where they actually wanted some sweets.

*I don't know...it may not be according to the idea with the concept, but rather than giving away magic bags [bags with unknown content to the consumer], they could list the items available online. The consumers can then pick...four items in a bag to an affordable price, rather than having 10 items in a bag and end up throwing half of them. [Male student, 29 years, on experiences from TGTG]*

But while most found the TGTG concept more concerned about the producers' interests and was neither an affordable economic solution nor promoting more healthy foods or less personal food waste, 1 student had many positive experiences and personal solutions to what others identified as problems. Rather than becoming annoyed by the large volume of bakery goods offered, she made sure the leftovers were stored in the freezer or given to friends. Furthermore, she had gained an experienced knowledge on what wholesalers to engage with, to find affordable offers with decent foods. She saved money on foods by engaging with the TGTG, and that it gave her opportunities to also have fresh food, like the bread of the day, by being willing to wait until the end of the day. She had continued using the TGTG based on the positive experiences, not only due to the feeling of really helping in reducing local food waste, but also motivated by personal interests.

*Previously I ate very much according to routines and habits: like I always ate the same for breakfast. But*

*now, when I suddenly end up with fish pudding in my bag, and I never used to have this previously, I come thinking; “Oh, ok, so this will be what I eat today”. So...I don’t know, but this makes me feel I actually eat more varied now. [Female student, 22 years, on experiences from TGTG]*

Compared with the TGTG app, the TotalCtrl app also included other functions, and provided the participants with an increased awareness of foods traits. Although they found the manual registration unnecessary, inconvenient, and time consuming, this step alerted them regarding expiry dates of foods and motivated them to better utilize foods within the household. The digital overview of the food stock and the provision of recipes within the app inspired some to buy less by impulse, to plan for optimal utilization of foods within the food stock, to keep an eye on expiry dates of the food stock, and to do more cooking. As such, providing accountability increased the potential for reducing food waste.

*You come to think about it more often. I haven’t reflected on it in a similar way previously, although I’ve been good at utilizing food I have available. But now, I more consciously think “I can use this left-over in other meals/menus too”. [Female student, 23 years, on experiences from TotalCtrl]*

*I feel like I’ve done more cooking in a way. When you get the recipes, and get some inspiration from it, you feel like doing it. [Female student, 23 years, on experiences from TotalCtrl]*

The students talked positively about the functionality in the TotalCtrl app, providing an overview of the food stock, the inspiration provided by menu suggestions, and the ease in transferring recipes to a shopping list. But this ironically also meant more cues for additional food shopping, which would be in contrast to reducing food waste.

*If I have milk in surplus that is due date, I’ve tried to find recipes based on milk. This makes me realize a need for other ingredients which I need to buy at the grocery store, and if I’m first going to the grocery store, then I could in reality buy something else and that I would not need the milk at all. [Female student, 22 years, on experiences from TotalCtrl]*

### Version 1.0

The experiences of the students in using the 2 different apps point to the many beneficial effects on awareness, accountability, positive experiences of reward by saving money or attaining cheap food, and personal contribution to reduce global food waste.

*It was really nice, with many recipes and related stuff. And that you can, with a simple click, just add all the ingredients from a recipe to your shopping list, it was really smart. [Male student, 29 years, on experiences from TotalCtrl]*

*I’ve experienced some eye openers. I have not thought of these thing previously....like if a thing is due date. (-) Now, living alone, I need to keep an eye on these things, and this project has contributed to increased*

*awareness. Yes, now I check due dates maybe once per week, so that I’m able to make plans. [Female student, 23 years, on experiences from TotalCtrl]*

However, several shortcomings and a potential for improvements of the apps were identified. The lack of consumer focus by the TGTG, mainly about reducing food waste at the producer level, was mentioned by most. The TotalCtrl, by contrast, was better suited for the consumer needs and included more functions. But one clear limitation was the lack of prize information, and as such, no function could help consumers in getting control over the budget or expenses. Furthermore, a small database and a slow development and update process were mentioned. The function with the barcode scanning first brought excitement, being automatic and quick, but later caused frustrations. These experiences also resulted in some ideas for improvement.

*The app is so...manually. When I tried to scan an item, the database was not big enough and it did not recognize the item, so therefor I had to type it in again, and then I also had to type in about expiration date, and sometimes you pick up an orange, and the...the item don’t have any given expiration date, right? [Female student, 27 years, on experiences from TotalCtrl]*

*It did not give me any good overview. I think it took too much time to register (manually) in the app-fridge. And then I had to make sure I updated it....like if I ate something, then I had to update that it’s no longer in the fridge. Just like that – it just became too much work, by something that should be very simple. [Female student, 22 years, on experiences from TotalCtrl]*

*I think maybe what could have made the function [barcode scanning] even better, would be that the scans I make of new items, like a box of macaroni, will automatically enter a database. Then when the next person scans this same barcode, there will appear notice that I know something about it. Yes, so everyone in a way helps to expand the data base with different items. [Male student, 29 years, on experiences from TotalCtrl]*

*If you have registered what you have in the fridge and in the freezer, and if you find a recipe in the app, then it automatically knows that you have four items from the recipe in the fridge, so that it does not add them to the shopping list, for example. [Male student, 29 years, on experiences from TotalCtrl]*

There are still some challenges that are difficult to resolve when it comes to the app initiatives for reducing personal food waste. Potentially this can be solved by adding a reminder function that notifies if food items have not been in use after being placed in the food stock.

*I thought the app would be smart and estimate time left for an item for me, but it did not. Like, I bought one kilo of avocados and left them in the fridge, and I did not remember about them, and they got rotten. So, I don’t think its really helpful in the case of*

reducing food waste. [Female student, 27 years, on experiences from TotalCtrl]

**Effects From Use of Digital Apps**

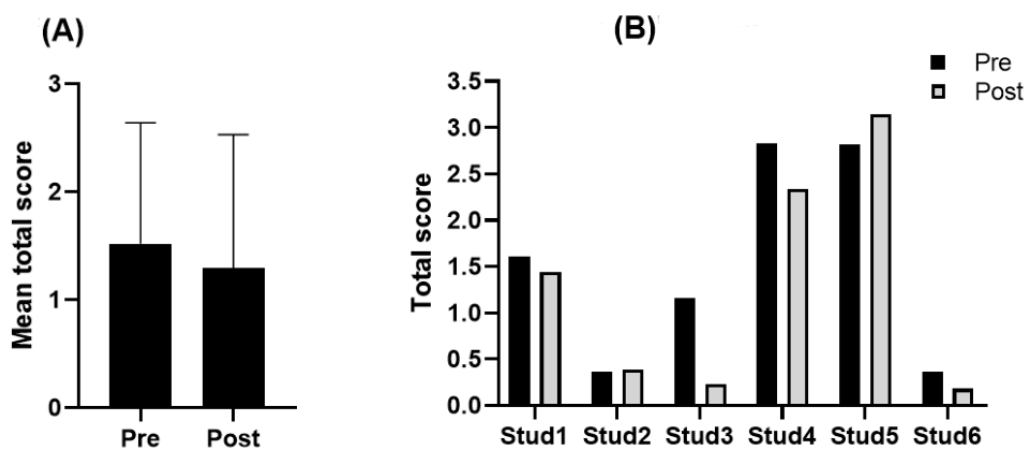
There was no statistically significant change in the healthy diet index score in the group, with a mean change of -0.24 (SD 0.43;  $P=.24$ ) from pre- to posttrial (Figure 2A). The individual changes point toward some different personal experiences (Figure 2B), but overall with small effects. There were few students consuming fish, at least on a regular basis, and among the 3 students reporting regular intake of fruits and vegetables, portions of 1-3 per day were most typical. While 2 participants were concerned about doing their own cooking, the sample in

general was characterized by high consumption of take-away foods.

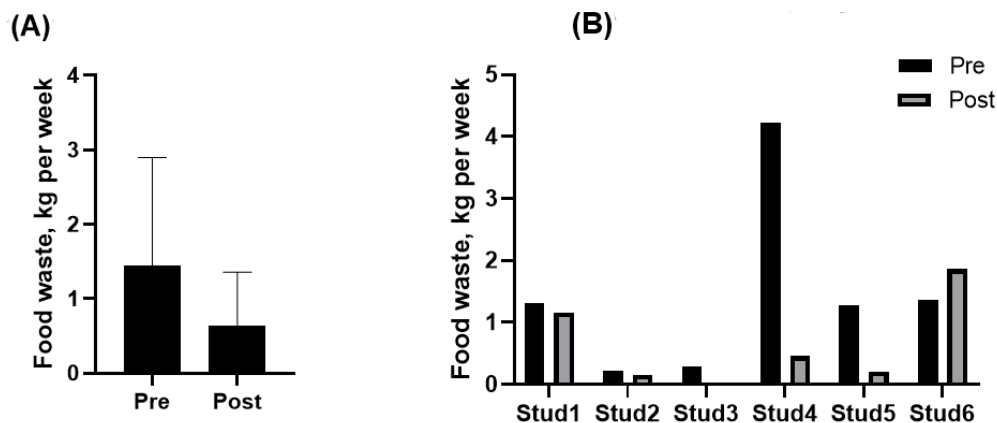
There was no significant reduction in food waste in the whole group of students (Figure 3A): mean change 0.81 (SD 1.5) kg ( $P=.13$ ). One student specifically experienced a large positive effect on food waste reduction during the trial, but the remaining students revealed negligible effects (Figure 3B).

There were no statistically significant differences in food expenses from before to after the app trials (Figure 4A): mean change 47.5 (SD 416.9) NOK or US \$4.8 (SD US \$42.5;  $P=.39$ ). The lack of change was seen as a result from half increasing their spending and half decreasing their spending (Figure 4B).

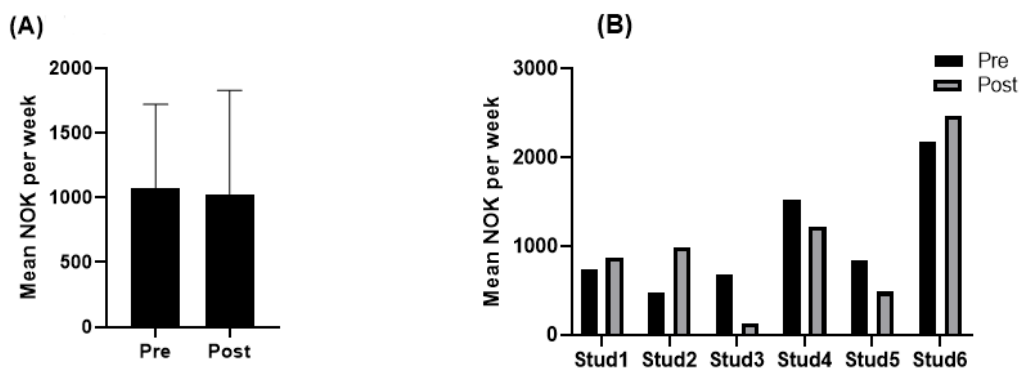
**Figure 2.** The healthy diet index (HDI). Panel A illustrates the group mean (StD) total score in HDI (ranging from 0-5, with higher scoring indicating more healthy diets) before (pre) and after (post) trials with the two digital apps; and panel B illustrates the individual changes in HDI per student. Stud: student.



**Figure 3.** The mean food waste before and after app-trials. Panel A illustrates the group mean (StD) total food waste before (pre) and after (post) trials with the two digital apps; and panel B illustrates the individual changes per student. Stud: student.



**Figure 4.** The mean economic expenses for food before and after app-trials. Panel A illustrates the group mean (StD) total expenses before (pre) and after (post) trials with the two digital apps; and panel B illustrates the individual changes per student. Stud: student.



## Discussion

### Principal Findings

This pilot intervention study evaluated the experiences of using smartphone apps designed to reduce food waste and financial food spending, and whether such apps could result in measurable changes, including an improved diet quality, among students. The main findings were that the apps increased awareness on food waste, but that they neither reduced food waste or financial expenses on food, nor improved healthy eating. Experiences from using the apps point toward a need to increase automatic operations and reminder functions, and increased attention to the needs of the personal consumer.

As much as students are aware of global climate issues, and truly engage in improving these, the immediate need to safeguard personal health and financial situation seems to overshadow the commitment to the environment. This does not imply an ignorance of the global perspective, rather they need to cope with their own situation before finding the capacity to focus on a broader perspective. With apps promoted to embrace both perspectives, there is a low barrier to engage. However, as long as the apps are more concerned about the producers' interests than the consumers' interests, the engagement and interest from the consumer may be short lived.

The affordable offers in these less-food-waste apps may actually conquer the objective of gaining control over personal health, which may be seen as a consequence when apps are primarily designed to match the interests of the producers and wholesalers. Young adults, such as students, are in situations that may cause more awareness about the relationship between behavior and health. This may simply be trying to live on their own for the first time, establishing a personal lifestyle behavior, or because moving to new places means large changes in (food) culture. If there is an increased awareness on how foods impact our health, commercial initiatives on reducing food waste may actually be counterproductive, specifically as long as the apps are predominantly designed for the producers' interests.

One important finding from this study is what emerges as the main determinant for efforts in reducing food waste, whether it is on the personal level or the producer's level. Although concerns on the environmental impact of food waste are real, the motivation and willingness to engage depend on the

experienced *economic* outcomes. While the apps are designed from the producer's perspective, that is, reducing food waste in store by offering last-minute offers, thus ensuring some income for the producer, the food waste in a wider perspective is not reduced. The customer may first be motivated by the initiative from these apps; users can contribute to less food waste while having edibles at a reduced price. However, when the edibles are foods they originally did not want or need, they realize that personal money is spent on unnecessary expenses and food waste is simply replaced from the producer level to the consumer level. Less food waste does not seem to be realized based on pure idealism, but must include an individual profit impacting individual resources if the parts are to engage.

### Comparison With Prior Work

Our findings, pointing to a need for personal reward, have previously been highlighted as important for successful behavioral changes and from app use specifically [17]. The experience of personal reward, being financial or other forms of profit (eg, attaining goods or improving health) after making personal efforts, is specifically reinforcing for the motivation to engage in a specific behavior. Self-monitoring is regarded as one of the most effective measures for behavioral change, as it brings alertness on effects from personal behavior, thereby enhancing commitment [17,25]. The feeling of reward is easily attainable when the efforts have been registered and measured, and such apps may aid in performing self-monitoring.

The students in this study had different experiences from the 2 apps in the trials, but most agreed that manual work and complex tasks reduced their interest to engage with the app. The lack of long-term compliance in the use of apps has previously been highlighted, and too much manual work was mentioned as one important limitation [26]. Some also mentioned the desire for more personalized content, like an opportunity to register specific dietary considerations to have more personalized offers and content in the apps. Besides the preference for automatic functions and personalized content, there was the need for reminders (eg, for the content in food stock getting close to the expiry date). These findings confirm previous literature reviews on the necessary functions of apps [17]. The apps did increase the awareness on global food waste, self-monitoring of food consumption, food waste registration, insight into the potential for health improvements, economical savings, and personal responsibility for reducing food waste. This points to the



potential for many personal and global beneficial outcomes, although the interests for engagement with such apps first need to be triggered. Underlining previous findings, education about consequences of wasting food and experiencing a joint commitment from the society may be important triggers [11]. But as previously reported, busy periods, such as those during examinations, experienced by students mean less motivation to engage, if reducing food waste means extra personal efforts (eg, manual work in the apps, needing to visit the stores at specific times to have the affordable offers, or cooking) [11,12].

### Limitations

The strengths of this pilot study are the use of weighed food and waste records before and after the trials, and the precise registrations of expenses by registration of all receipts. Further ensuring the credibility and generalization of the current findings is the inclusion of students of both sexes, from different study programs, and with different experiences in food registration and food waste engagement. The study also contributes with data on the content, usability, and efficacy of commercial apps, thus informing professionals in making more targeting apps [22].

Limiting the generalizability of our findings are the small group of participants and the use of only 2 apps. Nevertheless, referring to a previous review of available apps designed to reduce food waste, only 11 were identified, and findings on user satisfaction were in line with those currently reported [17]. Importantly, the financial reward that the students received from participating

in this study may have shaped the outcome of this study; rewards such as financial savings are important in motivating individuals to achieve reduction in food waste. Still, financial initiatives are commonly used as recruitment strategy in studies, and students are specifically concerned about their restricted economy, which naturally results in the preoccupation of or priority of saving money. Finally, of specific importance relating to the lack of quantifiable effects from this trial is the fact that students were facing their examination period during the posttrial measurements. Hence, most students said they did not have time to engage with app offers and recommendations, or do cooking or think of economical savings, and as such ate much take-away and on-the-go foods.

### Conclusion

Our findings suggest that apps designed to reduce food waste must combine the personal interests by the consumer (ie, economy, health, not too complex or detailed functions, and less manual operations) with the interest of the producers/wholesaler (economy). A frequent update on foods within the app database, inclusion of new providers, and maintenance and development of operating functions are important efforts from the app providers. Embracing these needs and perspective may better contribute to a total, global enhanced utilization of edible resources, and as such reduces food waste. Further prospective studies need to be conducted, preferably with different population samples, to examine whether apps designed to reduce food waste truly do have any beneficial effects.

### Acknowledgments

This trial has been initiated and sponsored by TotalCtrl AS with funding from the Norwegian Research Funding and Regional Qualification Project (id-number 323197). The authors received funding to participate in this project.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Supplementary data; interview guide referred to in the method-paragraph.

[PDF File (Adobe PDF File), 95 KB - formative\_v6i9e38520\_app1.pdf ]

### References

1. National Academies of Sciences, Engineering, and Medicine. A National Strategy to Reduce Food Waste at the Consumer Level. Washington, D.C: The National Academies Press; 2020:1-322.
2. Stenmarck Å, Jensen C, Quested T, Moates G. Estimates of European food waste levels: March 2016, FUSIONS project. Stockholm, Sweden: IVL Swedish Environmental Research Institute; Mar 2016.
3. Morone P, Koutinas A, Gathergood N, Arshadi M, Matharu A. Food waste: Challenges and opportunities for enhancing the emerging bio-economy. *Journal of Cleaner Production* 2019 Jun;221:10-16. [doi: [10.1016/j.jclepro.2019.02.258](https://doi.org/10.1016/j.jclepro.2019.02.258)]
4. Aamir M, Ahmad H, Javaid Q, Hasan SM. Waste Not, Want Not: A Case Study on Food Waste in Restaurants of Lahore, Pakistan. *Journal of Food Products Marketing* 2018 Jun 26;24(5):591-610. [doi: [10.1080/10454446.2018.1472695](https://doi.org/10.1080/10454446.2018.1472695)]
5. Bravi L, Francioni B, Murmura F, Savelli E. Factors affecting household food waste among young consumers and actions to prevent it. A comparison among UK, Spain and Italy. *Resources, Conservation and Recycling* 2020 Feb;153:104586. [doi: [10.1016/j.resconrec.2019.104586](https://doi.org/10.1016/j.resconrec.2019.104586)]
6. Thyberg KL, Tonjes DJ. Drivers of food waste and their implications for sustainable policy development. *Resources, Conservation and Recycling* 2016 Jan;106:110-123 [FREE Full text] [doi: [10.1016/j.resconrec.2015.11.016](https://doi.org/10.1016/j.resconrec.2015.11.016)]
7. European Union. European Parliament resolution of 16 May 2017 on initiative on resource efficiency: reducing food waste, improving food safety (2016/2223(INI)). *Official Journal of the European Union* 2018 Aug 30:25-43 [FREE Full text]

8. European Economic and Social Committee, Gheoldus M, Jan O, O'Connor C. Comparative study on EU Member States' legislations and practices on food donations: final report. European Economic and Social Committee. 2014. URL: <https://op.europa.eu/en/publication-detail/-/publication/149117ed-8eae-49c7-8369-1b968c800cdf/language-en> [accessed 2022-08-23]
9. Quested T, Parry A, Easteal S, Swannell R. Food and drink waste from households in the UK. *Nutrition Bulletin* 2011;36:460-467. [doi: [10.1111/j.1467-3010.2011.01924.x](https://doi.org/10.1111/j.1467-3010.2011.01924.x)]
10. Laurenti R, Moberg Å, Stenmarck Å. Calculating the pre-consumer waste footprint: A screening study of 10 selected products. *Waste Manag Res* 2017 Jan;35(1):65-78. [doi: [10.1177/0734242X16675686](https://doi.org/10.1177/0734242X16675686)] [Medline: [27837187](https://pubmed.ncbi.nlm.nih.gov/27837187/)]
11. Attiq S, Chau KY, Bashir S, Habib MD, Azam RI, Wong W. Sustainability of Household Food Waste Reduction: A Fresh Insight on Youth's Emotional and Cognitive Behaviors. *Int J Environ Res Public Health* 2021 Jun 30;18(13):7013 [FREE Full text] [doi: [10.3390/ijerph18137013](https://doi.org/10.3390/ijerph18137013)] [Medline: [34209149](https://pubmed.ncbi.nlm.nih.gov/34209149/)]
12. Singhirunnusorn W, Donlakorn K, Kaewhanin W. Contextual Factors Influencing Household Recycling Behaviours: A Case of Waste Bank Project in Mahasarakham Municipality. *Procedia - Social and Behavioral Sciences* 2012;36:688-697. [doi: [10.1016/j.sbspro.2012.03.075](https://doi.org/10.1016/j.sbspro.2012.03.075)]
13. Visschers V, Wickli N, Siegrist M. Sorting out food waste behaviour: A survey on the motivators and barriers of self-reported amounts of food waste in households. *Journal of Environmental Psychology* 2016 Mar;45:66-78 [FREE Full text] [doi: [10.1016/j.jenvp.2015.11.007](https://doi.org/10.1016/j.jenvp.2015.11.007)]
14. Burlea-Schiopoiu A, Ogarca RF, Barbu CM, Craciun L, Baloi IC, Mihai LS. The impact of COVID-19 pandemic on food waste behaviour of young people. *J Clean Prod* 2021 Apr 20;294:126333 [FREE Full text] [doi: [10.1016/j.jclepro.2021.126333](https://doi.org/10.1016/j.jclepro.2021.126333)] [Medline: [34720458](https://pubmed.ncbi.nlm.nih.gov/34720458/)]
15. Knezevic B, Kurnoga N, Anic I. Typology of university students regarding attitudes towards food waste. *BFJ* 2019 Sep 17;121(11):2578-2591. [doi: [10.1108/bfj-05-2018-0316](https://doi.org/10.1108/bfj-05-2018-0316)]
16. Statista. Smartphone ownership in the US by age 2021. Statista. Hamburg, Germany: Statista GmbH; 2022. URL: <https://www.statista.com/statistics/489255/percentage-of-us-smartphone-owners-by-age-group/> [accessed 2022-08-23]
17. Nkwo M, Suruliraj B, Orji R. Persuasive Apps for Sustainable Waste Management: A Comparative Systematic Evaluation of Behavior Change Strategies and State-of-the-Art. *Front Artif Intell* 2021;4:748454 [FREE Full text] [doi: [10.3389/frai.2021.748454](https://doi.org/10.3389/frai.2021.748454)] [Medline: [34957392](https://pubmed.ncbi.nlm.nih.gov/34957392/)]
18. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. *J Med Internet Res* 2013 Apr 18;15(4):e86 [FREE Full text] [doi: [10.2196/jmir.2583](https://doi.org/10.2196/jmir.2583)] [Medline: [23598614](https://pubmed.ncbi.nlm.nih.gov/23598614/)]
19. Fogg BJ. *Persuasive Technology. Using Computers to Change What We Think and Do* (1st Edition). Amsterdam, The Netherlands: Elsevier/Morgan Kaufmann; Dec 16, 2002:1-312.
20. World Health Organization (WHO). World Health Organization, Fact sheets: Healthy Diet. WHO. Geneva, Switzerland: World Health Organization; 2020. URL: <https://www.who.int/news-room/fact-sheets/detail/healthy-diet> [accessed 2022-08-23]
21. World Health Organization (WHO). World Health Organization, Fact sheet: Noncommunicable diseases. WHO. Geneva, Switzerland: World Health Organization; 2017. URL: <http://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases> [accessed 2022-08-23]
22. 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](https://doi.org/10.2196/mhealth.8758)] [Medline: [29254914](https://pubmed.ncbi.nlm.nih.gov/29254914/)]
23. Nordic Council of Ministers. *Nordic Nutrition Recommendations 2012 - Integrating Nutrition and Physical Activity* (5th Edition). Copenhagen, Denmark: The Nordic Council of Ministers; 2012.
24. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp0630a](https://doi.org/10.1191/1478088706qp0630a)]
25. Morrow A, Walker K, Calder-MacPhee N, Ozakinci G. The active ingredients of physical activity and / or dietary workplace-based interventions to achieve weight loss in overweight and obese healthcare staff: a systematic review. *J Behav Med* 2022 Jun;45(3):331-349. [doi: [10.1007/s10865-021-00279-x](https://doi.org/10.1007/s10865-021-00279-x)] [Medline: [35132501](https://pubmed.ncbi.nlm.nih.gov/35132501/)]
26. Ögmundsdóttir Michelsen H, Sjölin I, Bäck M, Gonzalez Garcia M, Olsson A, Sandberg C, et al. Effect of a Lifestyle-Focused Web-Based Application on Risk Factor Management in Patients Who Have Had a Myocardial Infarction: Randomized Controlled Trial. *J Med Internet Res* 2022 Mar 31;24(3):e25224 [FREE Full text] [doi: [10.2196/25224](https://doi.org/10.2196/25224)] [Medline: [35357316](https://pubmed.ncbi.nlm.nih.gov/35357316/)]

---

## Abbreviations

**TGTG:** Too-Good-To-Go

---

*Edited by A Mavragani; submitted 05.04.22; peer-reviewed by R Davis-Martin, E Hohman; comments to author 20.06.22; revised version received 01.07.22; accepted 14.07.22; published 02.09.22.*

*Please cite as:*

*Mathisen TF, Johansen FR*

*The Impact of Smartphone Apps Designed to Reduce Food Waste on Improving Healthy Eating, Financial Expenses and Personal Food Waste: Crossover Pilot Intervention Trial Studying Students' User Experiences*

*JMIR Form Res 2022;6(9):e38520*

*URL: <https://formative.jmir.org/2022/9/e38520>*

*doi: [10.2196/38520](https://doi.org/10.2196/38520)*

*PMID: [36053667](https://pubmed.ncbi.nlm.nih.gov/36053667/)*

©Therese Fostervold Mathisen, Frode Ramstad Johansen. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 02.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# The Use of Smartphone Serious Gaming Apps in the Treatment of Substance Use Disorders: Observational Study on Feasibility and Acceptability

Thelma Schilt<sup>1,2,3</sup>, PhD; Elvira Sharine Ruijter<sup>1,3,4</sup>, MD; Nikky Godeschalk<sup>1,2,3</sup>, MSc; Marit van Haaster<sup>1,2</sup>, MSc; Anna E Goudriaan<sup>1,3,5</sup>, Prof Dr

<sup>1</sup>Jellinek, Amsterdam Institute for Addiction Research, Arkin Mental Health Care, Amsterdam, Netherlands

<sup>2</sup>Department of Psychology, Brain&Cognition, University of Amsterdam, Amsterdam, Netherlands

<sup>3</sup>Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands

<sup>4</sup>VU-University Medical Center, School of Medical Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

<sup>5</sup>Department of Research and Quality of Care, Arkin Mental Health Care, Amsterdam, Netherlands

**Corresponding Author:**

Thelma Schilt, PhD

Jellinek, Amsterdam Institute for Addiction Research

Arkin Mental Health Care

Jacob Obrechtstraat 92

Amsterdam, 1071 KR

Netherlands

Phone: 31 88 505 1220

Email: [thelma.schilt@jellinek.nl](mailto:thelma.schilt@jellinek.nl)

## Abstract

**Background:** Addiction is a worldwide problem with major health complications. Despite intensive treatment, relapse rates remain high. The prevalence of cognitive impairment is high in patients with substance use disorders (SUDs) and is associated with treatment dropout and relapse. Evidence indicates that cognitive function training in persons with SUDs may support treatment. Therefore, the use of web-based tools to test and train cognitive functions is of increasing interest.

**Objective:** The goal of this study was to determine the feasibility and acceptability of a serious gaming smartphone app to test and train cognitive functions in addition to the treatment of SUDs.

**Methods:** A prospective observational study was conducted with 229 patients seeking addiction treatment. The patients were offered 2 smartphone apps in addition to regular care: MyCognition Quotient (MyCQ) assessed cognitive functions and AquaSnap trained these functions. The feasibility was determined based on acceptance rates. The acceptability of the smartphone apps was qualitatively analyzed based on the answers to a questionnaire. Patient characteristics were compared between patients who played and did not play smartphone games. Explorative correlation analyses were performed between the playing time and cognitive assessment scores.

**Results:** Of the 229 patients who were offered the apps, 110 completed the MyCQ assessment, and 59 started playing AquaSnap, yielding acceptance rates of 48.0% and 25.8%, respectively. The group that completed the MyCQ assessment was significantly more educated than the group that did not download the apps ( $\chi^2_2=7.3$ ;  $P=.03$ ). The education level did not differ significantly between the group that played AquaSnap and the group that did not ( $P=.06$ ). There were relatively more women in the AquaSnap playing group than in the nonplaying group ( $\chi^2_1=6.5$ ;  $P=.01$ ). The groups did not differ in terms of age, substance use, treatment setting, mood, or quality of life. With respect to acceptability, 83% (38/46) of the patients who filled out the questionnaire enjoyed taking the MyCQ measurement, whereas 41% (14/34) enjoyed playing the AquaSnap game. Furthermore, 76% (35/46) and 68% (23/34) rated the apps MyCQ and AquaSnap, respectively, as easy. More playing minutes was associated with decreased working memory reaction time and executive functioning accuracy.

**Conclusions:** Our study showed that the use of a smartphone app for cognitive assessment in patients with SUDs who are interested and highly educated is feasible and acceptable for the subgroup that was asked to fill out a perception questionnaire. However, the use of a smartphone app for cognitive training was less feasible for this group of patients. Improvement of the training application and enhancement of the motivation of clients are needed. Despite these limitations, the present results provide

support for future research investigating the use of smartphone apps for cognitive assessment and training in relation to the treatment of SUDs.

(*JMIR Form Res* 2022;6(9):e34159) doi:[10.2196/34159](https://doi.org/10.2196/34159)

## KEYWORDS

addiction care; mobile phone; cognitive training; neurocognition; mental health; mobile health; digital applications; health applications; smartphone; cognitive assessment

## Introduction

### Background

Worldwide, approximately 269 million people used drugs in 2018 (World Drug Report 2020, United Nations Office on Drugs and Crime), and among them, an estimated 35.6 million people had substance use disorders (SUDs). Approximately 12.5%, that is, 4.5 million people with SUDs received treatment. A common intervention is psychological therapy, such as cognitive behavioral therapy, motivational interviewing, and contingency management [1,2], often in combination with medications. Although there is evidence that these interventions are effective [1,2], relapse rates remain high [3], and approximately half of the patients treated for SUDs relapse within 1 year after treatment [4]. Dropout is a predictor of relapse [5], with nearly one-third of the patients dropping out of psychosocial SUDs treatment [6]. One of the risk factors for dropout from treatment is the presence of cognitive deficits [7]. Moreover, cognitive impairment is associated with an increased chance of addiction relapse [8,9], even after recovery from SUDs [10]. Bruijnen et al [11] showed that cognitive impairment was present in 31% of patients with SUDs who presented for treatment. Substance use can impair executive functioning, such as decision-making, flexibility, planning, and inhibition [12], and in addition, weaker executive functions have been related to the development of addictive disorders (eg, [13,14]). Depending on the substance abused, specific cognitive functions may be affected [12,15,16]. Moreover, different cognitive impairments predict relapse, and training could be aimed at improving those cognitive functions [10].

Early detection of cognitive impairment in addiction care is of clinical importance, because cognitive training can be provided in time to optimize addiction treatment. A combination of behavioral therapy and cognitive training may improve treatment outcomes [17]. Preliminary evidence shows promising results for cognitive training in addiction treatment [18]. A recent review article supported the usefulness of structured cognitive training programs in addition to conventional addiction treatment to improve cognitive performance in patients with SUDs; however, a limited number of studies have also evaluated SUD clinical outcomes, such as substance reduction or relapse prevention [9]. Improved cognitive functioning can lead to better social inclusion and support [19].

The content, methods, and applications of cognitive assessments vary widely, and the requirement for specialist supervision is time consuming and expensive. In the search for cheaper and more efficient methods, mobile apps are of increasing interest. Mobile apps have good accessibility and low costs and can be used to monitor and potentially improve mental health [20].

Previous research in adolescents who abused alcohol showed a significant improvement in behavioral control in adolescents who were trained with a serious game based on a stop-signal paradigm [21]. In addition, improvement in frontal cognitive functioning after training with a serious game was found in patients undergoing alcohol rehabilitation [22]. Another study in male Veterans with alcohol use disorder suggests that serious games that emphasize relapse prevention intervention techniques have positive effects on self-reported ratings of alcohol dependence, alcohol craving, and self-efficacy [23]. In a study of patients dependent on heroin, undergoing methadone maintenance treatment, a serious game was used in the treatment program, but because of the small sample size, it was uncertain whether the improvement at follow-up was due to the cognitive intervention [24].

### Objectives

A web-based tool that can be run on a smartphone, MyCognition Quotient (MyCQ), was developed to quickly and easily assess the broad cognitive status of patients. The app was validated in patients with obsessive-compulsive disorder, schizophrenia, and major depressive disorder [25]. MyCQ is used in unison with a web-based training application AquaSnap, where it tracks progress and determines which cognitive domains require the most training. Recently, the beneficial effects of the training application AquaSnap on the perception of subjective cognitive functioning were found in a group of patients with breast cancer [26]. It is of great value to investigate whether these 2 applications are useful in patients with SUDs. As advised by the National Institute of Health Research, this study focused on the feasibility of such a study [27]. As these smartphone apps have not yet been studied for the treatment of addiction, their compliance and acceptability are not known. Therefore, this study examined the feasibility and acceptability of these apps in a group of patients with SUDs in an addiction treatment setting. The association between playing time and cognitive functioning was explored in patients who used the cognitive training application.

## Methods

### Ethics Approval

This study was categorized as not subject to the Medical Research Involving Humans Subjects Act (*Wet Medisch-wetenschappelijk Onderzoek met mensen*) by the Medical Ethics Committee of the Amsterdam University Medical Centers. Informed consent was obtained from all participants before the start of the study, and the study was performed in accordance with the Declaration of Helsinki.

## Study Settings and Design

In a prospective study, the feasibility of implementing serious games in addition to addiction care was investigated. The target population was patients seeking treatment in an addiction-treatment center, Jellinek, Amsterdam, comprising an outpatient and inpatient facility, including a detoxification facility. Treatment was voluntary, and the patients and participants could stop treatment at any time. Convenience sampling was used to recruit patients.

## Participants

Participants were recruited between April 2019 and June 2020. Patients starting cognitive behavioral therapy or acceptance and commitment therapy at the detoxification, inpatient, or outpatient units were informed regarding the study through leaflets and presentations. Interested patients could participate in the study if they fulfilled the criteria of substance use or gambling disorder according to the Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition, were aged between 18 and 75 years, could read Dutch or English, and were capable of using a smartphone. Patients were excluded if they were addicted to gaming or gambling on their smartphones or the internet.

## Measures

### Patient Characteristics

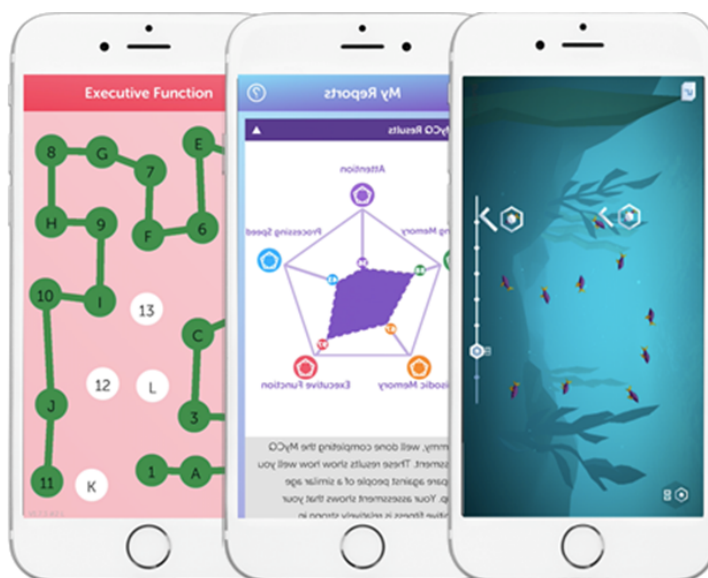
Demographic and clinical characteristics of the participants were collected (age, gender, highest level of education, substance use, and current addiction treatment):

- Measurements in the Addictions for Triage and Evaluation [28] was administered to collect information on substance use and SUDs in the past month and lifetime.
- The 21-item Depression Anxiety Stress Scale (DASS) was used to measure depression, anxiety, and stress [29-31]. The DASS-21 is a self-report questionnaire consisting of 21 items, with 7 items per subscale: depression, anxiety, and stress. Patients were asked to score every item on a scale from 0 (did not apply to me at all) to 3 (applied to me very much). Sum scores were computed by summing the scores on the items per scale and multiplying them by a factor of 2. Sum scores for the total DASS range between 0 and 120, and those for each of the subscales may range between 0 and 42.
- The Manchester Short Assessment of Quality of Life (MANSA) questionnaire was administered to measure quality of life [32]. The questionnaire contained 12 questions about satisfaction with life as a whole, including occupational status (eg, job and sheltered employment), financial situation, number and quality of friendships, leisure activities, accommodation, personal safety, people with whom the patient lives (or living alone), sex life, relationship with family, physical health, and mental health. Satisfaction was rated on a 7-point scale (1=negative extreme; 7=positive extreme).

### Intervention

In this study, 2 smartphone apps developed by MyCognition were used. These apps are available on the web and can be used at home without the help of a trained supervisor. Figure 1 [33] shows a screenshot of these apps.

**Figure 1.** Screenshots of one of the tests and overview page of the MyCognition Quotient neurocognitive assessment, with scores on the five domains assessed (attention, executive functions, episodic memory, and processing speed). On the right, a screenshot of one of the AquaSnap serious games [44].



The app MyCQ assesses cognitive functioning in 5 domains: attention, processing speed, working memory, episodic memory, and executive function. Every subtest starts with a practice trial. Attention is measured using a choice reaction time test, asking the participant to tap the screen when a red dot appears. The processing speed is measured by tapping the screen when a

stimulus appears. Working memory is tested using a 2-back task, where participants have to respond when a stimulus matches a picture 2 steps earlier in the sequence. Visual memory is tested by displaying a series of pictures to be remembered. Hereafter, the participant presses a button if they recognize the picture in a new list. Executive functioning is tested with a

trail-making task in which the participant has to alternately connect letters with numbers. Latency (reaction time) and accuracy (percentage correct) scores are recorded for each domain.

The second app (AquaSnap) is a videogame that improves cognitive functions and targets the 5 cognitive domains assessed by MyCQ. The training program adapts to the individual MyCQ assessment scores and AquaSnap training performance. The app consists of 7 games that each train one or more cognitive domains. In the games, the player is a submarine that dives underwater and can discover different parts of the ocean. Taking the best pictures of fish and completing missions provides the player experience and currency, which can be used to discover new areas. A detailed description can be found in the study by Domen et al [25].

## Outcome Measures

### Feasibility of Engagement

In this study, the apps were considered feasible if 24.8% (57/229) of the recruited patients completed the MyCQ assessment and, of these, 60% (34/57) managed to play AquaSnap for a minimum of 15 minutes [34,35].

### Acceptability of the Apps

A questionnaire was administered to assess the acceptability of the 2 smartphone apps. Questions about ease of use and likeability were rated on a 5-point Likert scale, where 1=totally disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, and 5=totally agree.

The use of the apps was assumed to be acceptable if at least 30% (14/46 and 10/34, respectively) of the participants who filled out the questionnaire rated the apps positively in terms of ease of use and likeability, that is, if they rated it with a score of 4 (agree) or 5 (totally agree) on a 5-point Likert scale [35].

### Procedure

Participants were instructed to download the apps *MyCQ* and *AquaSnap* from the App or Play Store on their smartphone—with an information leaflet and email sent to them. Logging into the app required a log-in code that was provided by the therapist. If participants had difficulties downloading the app, the therapist helped them with it in a treatment session. All participants were allocated a participant number to be used for the app during its download, and no participant-related identifiers were captured in electronic data files. Electronic data from the smartphone app were synchronized to a password-protected cloud database. The playing time of AquaSnap and the performance of the MyCQ tests were tracked for 6 weeks. In a separate password-protected electronic data file, participant numbers were connected to electronic patient file numbers to be able to contact the participants during the study period. The participants were instructed to start with the MyCQ assessment because the AquaSnap game adapted to the MyCQ test results. Every week, if little or no activity was observed, the participant was called and, if reached, motivated to use the apps. It was advised to play the AquaSnap game for 15 minutes daily, as in previous studies, longer playing times

were advised, but shorter times were played [26,36,37]. To stimulate playing the game, the participants who had played AquaSnap for >225 minutes received a €15 (US \$15.30) bol.com gift card. To assess acceptability, the completion of at least one MyCQ assessment and a minimum playing time of 15 minutes on AquaSnap was set to ensure that participants had some experience in answering a questionnaire about the use of the apps. In total, 41.8% (46/110) of participants who completed an assessment and 58% (34/59) who played AquaSnap were asked to answer a questionnaire about the use of the apps, which earned them a €10 (US \$10.20) gift voucher. The participants were free to stop playing the game at any moment. Because the primary aim was to assess feasibility and acceptability, no additional minimum time, besides the 15-minute limit, was used in the exploratory analyses for the effects of playing AquaSnap on cognition. Assessments of MyCQ, DASS, MANSA, and Measurements in the Addictions for Triage and Evaluation were performed at the start of the treatment and after 6 weeks.

### Analyses

Data were analyzed using SPSS Statistics (version 26; IBM Corp), and the statistical significance was set at .05. The group of participants who performed at least one assessment on their smartphone and the group of participants who did not use the apps at all were compared at baseline based on age, gender, education level, substance use, DASS, MANSA, and treatment type. A *t* test (2-tailed) was performed for continuous variables and a chi-square test for categorical variables (or Mann-Whitney *U* test or Fisher exact test, respectively, as appropriate). Within the group of participants who performed at least one MyCQ assessment, the same comparisons were made between participants who played AquaSnap for at least 15 minutes and those who did not. Mann-Whitney *U* tests were performed to compare the baseline MyCQ scores among the groups. For the MyCQ, we considered both latency (milliseconds) and accuracy (% true) for each of the 5 cognitive domains; therefore, the *P* value was set at .05/2=.03.

The acceptability of the smartphone apps was qualitatively analyzed based on the answers to the questionnaires.

Explorative Spearman correlation analyses were performed between the number of playing minutes on AquaSnap and the change in scores between the first and second MyCQ measurements. For latency, the change scores were calculated by subtracting the latency (speed) in milliseconds at follow-up from the latency in milliseconds at the start. For accuracy, the change scores were calculated by subtracting the accuracy in % true at the start from the % true at the follow-up.

## Results

### Overview

A total of 229 patients (151/229, 65.9% men) seeking addiction treatment for SUD were interested in participating in the study and received a log-in code to use the MyCQ and AquaSnap apps. The mean patient age was 42 (SD 12.7; range 19.4-74.8) years. The patient characteristics are presented in Table 1.

**Table 1.** Patient characteristics in the total sample and groups that used and did not use the app MyCognition Quotient (MyCQ).

Patient characteristics	Total	MyCQ	Non-MyCQ	<i>P</i> value <sup>a</sup>	Test
<b>Gender, n (%)</b>	229 <sup>b</sup> (100)	110 (100)	119 (100)	.58	Chi-square
Male	152 (66.4)	75 (68.2)	77 (64.7)		
Female	77 (33.6)	35 (31.8)	42 (35.3)		
Age (years), mean (SD)	42.0 (12.7)	43.7 (12.9)	40.4 (12.3)	.06	Mann-Whitney
<b>Level of education, n (%)</b>	174 (100)	87 (100)	87 (100)	.03 <sup>c</sup>	Chi-square
Low	20 (11.5)	5 (5.7)	15 (17.2)		
Average	63 (36.2)	30 (34.5)	33 (37.9)		
High	91 (52.3)	52 (59.8)	39 (44.8)		
<b>Major substance, n (%)</b>	217 (100)	105 (100)	112 (100)	.25	Chi-square
Alcohol	101 (46.5)	53 (50.5)	48 (42.8)		
Cannabis	47 (21.7)	20 (19)	27 (24.1)		
Cocaine	24 (11.1)	10 (9.5)	14 (12.5)		
Nicotine	15 (6.9)	8 (7.6)	7 (6.3)		
Stimulants	9 (4.1)	2 (1.9)	7 (6.3)		
Sedatives	7 (3.2)	2 (1.9)	5 (4.5)		
Other	6 (2.8)	5 (4.8)	1 (0.9)		
Gambling	8 (3.7)	5 (4.8)	3 (2.7)		
<b>Treatment setting, n (%)</b>	174 (100)	87 (100)	87 (100)	.45	Chi-square
Policlinic	96 (55.2)	51 (58.6)	45 (51.7)		
Daycare	41 (23.6)	17 (19.5)	24 (27.6)		
Clinic	37 (21.3)	19 (21.8)	18 (20.7)		
<b>DASS-t0<sup>d</sup>, mean (SD)</b>	149 (100)	75 (100)	74 (100)		<i>t</i> test (2-tailed)
Depression	15.8 (11.2)	17.4 (11.7)	14.2 (10.4)	.08	
Anxiety	10.4 (9.1)	10.6 (9.6)	10.2 (8.5)	.82	
Stress	15.9 (8.8)	16.5 (8.7)	15.2 (8.9)	.35	
Total	42.0 (25.8)	44.5 (26.9)	39.6 (24.7)	.25	
<b>MANSA-t0<sup>e</sup>, mean (SD)</b>	136 (100)	69 (100)	67 (100)		<i>t</i> test (2-tailed)
Total	51.8 (13.3)	51.6 (12.9)	52.1 (13.7)	.85	

<sup>a</sup>Between-group comparisons.

<sup>b</sup>Because of missing data, the n included is mentioned separately.

<sup>c</sup>Significant at .05 level

<sup>d</sup>DASS-t0: Depression Anxiety Stress Scale scores at baseline.

<sup>e</sup>MANSA-t0: Manchester Short Assessment of Quality of Life scores at baseline.

## Feasibility of Engagement

Of the 229 recruited patients, 110 completed the MyCQ assessment, providing an acceptance rate of 48%. Of these 110 MyCQ completers, 59 (53.6%) started playing AquaSnap, which was 25.8% (59/229) of the originally recruited patients.

## Comparison of the Participants Who Performed at Least One Assessment With Those Who Did Not Download the Apps

Of the patients who received a log-in code, 48% (110/229) downloaded the MyCQ app and completed at least one assessment. The group of 110 persons who completed the assessment did not differ in age, gender distribution, substance use, treatment setting, DASS, or MANSA scores from the 119 persons who did not download the apps (Table 1). The group



that completed the assessment was significantly more educated than the group that did not download the apps ( $\chi^2_2=7.3$ ;  $P=.03$ ).

### Comparison Within the MyCQ User Group Between AquaSnap Players and Nonplayers

In total, 59 participants played AquaSnap for a minimum of 15 (mean 194, SD 265, range 15-1032) minutes. They did not differ from the 51 participants who did not play AquaSnap in terms of age, substance use, treatment setting, DASS, or MANSA

scores (Table 2). There were significantly more women (25/59, 42%) in the AquaSnap group than in the non-AquaSnap group (10/51, 20%;  $\chi^2_1=6.5$ ;  $P=.01$ ). The educational level did not differ significantly between the AquaSnap and non-AquaSnap groups ( $\chi^2_2=5.6$ ;  $P=.06$ ). Baseline MyCQ scores did not differ between AquaSnap players and nonplayers (Table 3), except for episodic memory accuracy ( $U=1086$ ;  $P=.01$ ), which was better for AquaSnap players than for nonplayers.

**Table 2.** Patient characteristics of the participants who used the MyCognition Quotient app and per those who used and did not use the AquaSnap app.

Patient characteristics	Total	AquaSnap	Non-AquaSnap	<i>P</i> value <sup>a</sup>	Test
<b>Gender, n (%)</b>	110 <sup>b</sup> (100)	59 (100)	51 (100)	.01	Chi-square
Male	75 (68.2)	34 (57.6)	41 (80.4)		
Female	35 (31.8)	25 (42.4)	10 (19.6)		
Age (years), mean (SD)	43.7 (12.9)	44.8 (12.4)	42.6 (13.5)	.27	Mann-Whitney
<b>Level of education, n (%)</b>	87 (100)	46 (100)	41 (100)	.06	Chi-square
Low	5 (5.7)	4 (8.7)	1 (2.4)		
Average	30 (34.5)	11 (23.9)	19 (46.3)		
High	52 (59.8)	31 (67.4)	21 (51.2)		
<b>Major substance, n (%)</b>	105 (100)	57 (100)	48 (100)	.29	Chi-square
Alcohol	53 (50.5)	29 (50.9)	24 (50)		
Cannabis	20 (19)	12 (21.1)	8 (16.7)		
Cocaine	10 (9.5)	5 (8.8)	5 (10.4)		
Nicotine	8 (7.6)	6 (10.5)	2 (4.2)		
Stimulants	2 (1.9)	2 (3.5)	0 (0)		
Sedatives	2 (1.9)	0 (0)	2 (4.2)		
Other	5 (4.8)	2 (3.5)	3 (6.3)		
Gambling	5 (4.8)	1 (1.8)	4 (8.3)		
<b>Treatment setting, n (%)</b>	87 (100)	47 (100)	41 (100)	.40	Chi-square
Polyclinic	51 (58.6)	30 (65.2)	21 (51.2)		
Daycare	17 (19.5)	8 (17.4)	9 (22)		
Clinic	19 (21.8)	8 (17.4)	11 (26.8)		
<b>DASS-t0<sup>c</sup>, mean (SD)</b>	75 (100)	44 (100)	31 (100)		<i>t</i> test, 2-tailed
Depression	17.4 (11.7)	17.5 (10.8)	17.3 (13.1)	.95	
Anxiety	10.6 (9.6)	11.6 (10.4)	9.0 (8.3)	.25	
Stress	16.5 (8.7)	16.9 (8.8)	16.1 (8.6)	.70	
Total	44.5 (26.9)	46 (27.1)	42.4 (26.7)	.58	
<b>MANSA-t0<sup>d</sup>, mean (SD)</b>	69 (100)	40 (100)	29 (100)		<i>t</i> test, 2-tailed
Total	51.6 (12.9)	50.4 (12.6)	53.3 (13.5)	.35	

<sup>a</sup>Between-group comparisons.

<sup>b</sup>Because of missing data, the n included is mentioned separately.

<sup>c</sup>DASS-t0: Depression Anxiety Stress Scale scores at baseline.

<sup>d</sup>MANSA-t0: Manchester Short Assessment of Quality of Life scores at baseline.

**Table 3.** Mann-Whitney U test results for the comparison of baseline MyCognition Quotient (MyCQ) scores between participants who used and did not use AquaSnap.

MyCQ assessment	AquaSnap (n <sup>a</sup> =59), mean (SD)	Non-AquaSnap (n=50), mean (SD)	P value
Attention latency <sup>b</sup>	516 (106)	515 (139)	.89
Attention accuracy <sup>c</sup>	95 (10)	93 (17)	.69
Processing speed latency	362 (70)	356 (75)	.94
Processing speed accuracy	96 (5)	93 (17)	.22
Working memory latency	1312 (403)	1283 (444)	.58
Working memory accuracy	89 (14)	87 (15)	.41
Episodic memory latency	1128 (205)	1202 (428)	.56
Episodic memory accuracy	92 (7)	89 (8)	.01 <sup>d</sup>
Executive functioning latency	1417 (1298)	1716 (1535)	.16
Executive functioning accuracy	91 (14)	89 (16)	.21

<sup>a</sup>Owing to missing data for the subtest attention, n was 54 and 47, respectively.

<sup>b</sup>Latency in milliseconds.

<sup>c</sup>Accuracy in % true.

<sup>d</sup>This correlation was significant at the .03 level.

### Acceptability of the Apps

Of the 110 participants who performed the MyCQ assessment, 46 (41.8%) were asked to complete a questionnaire on the acceptability of the MyCQ assessment (Table 4). Of the 46 participants who completed the assessment, 38 (83%) enjoyed taking it, whereas only 3 (7%) disliked it. The majority (35/46,

76%) rated the app as easy. Furthermore, 74% (34/46) believed that the app provided insight into their brain functions. Half (23/46, 50%) of the participants who completed the questionnaire did not believe that the MyCQ assessments contributed to their addiction treatment and did not continue taking the measurements after finishing their treatment.

**Table 4.** Frequencies (%) perception questionnaire regarding serious gaming apps.

	Ratings <sup>a</sup> , n (%)					Value, mean (SD)
	1	2	3	4	5	
<b>MyCQ<sup>b</sup> assessment (n=46)</b>						
1. Did you like the MyCQ task?	1 (2)	2 (4)	5 (11)	28 (61)	10 (22)	4 (0.8)
2. Did the MyCQ task contribute to your addiction treatment?	13 (28)	10 (22)	14 (30)	7 (15)	2 (4)	2.5 (1.2)
3. Did the MyCQ task provide insight into your brain functions?	4 (9)	3 (7)	5 (11)	31 (67)	3 (7)	3.6 (1)
4. Would you continue with MyCQ after your treatment?	15 (33)	10 (22)	5 (11)	13 (28)	3 (7)	2.5 (1.4)
5. Was MyCQ easy in use?	1 (2)	6 (13)	4 (9)	20 (43)	15 (33)	3.9 (1)
<b>AquaSnap (n=34)</b>						
1. Did you like the AquaSnap game?	1 (3)	11 (32)	8 (24)	8 (24)	6 (18)	3 (1)
2. Do you think your brain functions are improved by playing AquaSnap?	4 (12)	7 (21)	7 (21)	14 (41)	2 (6)	2.8 (1.3)
3. Did AquaSnap contribute to your addiction treatment?	9 (27)	5 (15)	8 (24)	10 (29)	2 (6)	2.5 (1.3)
4. Did playing AquaSnap help you better manage your addiction?	11 (32)	11 (32)	7 (21)	4 (12)	1 (3)	2.1 (1.1)
5. Do you think the chance to relapse have diminished through AquaSnap?	13 (38)	7 (21)	10 (29)	3 (9)	1 (3)	2.1 (1.1)
6. Would you continue with AquaSnap after your treatment?	11 (32)	5 (15)	8 (24)	7 (21)	3 (9)	2.4 (1.4)
7. Did you find AquaSnap easy in use?	1 (3)	6 (18)	4 (12)	17 (50)	6 (18)	3.3 (1.3)

<sup>a</sup>1=totally disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, and 5=totally agree.

<sup>b</sup>MyCQ: MyCognition Quotient.

Of the 59 participants who played AquaSnap for at least 15 minutes, 34 (58%) completed a questionnaire on the acceptability of the AquaSnap app. Almost half (16/34, 47%) of the participants believed that playing the game improved their brain functions. The majority (23/34, 68%) rated the app as easy. Furthermore, 41% (14/34) of the participants enjoyed playing the game, whereas 35% (12/34) did not like it.

Moreover, 35% (12/34) of the participants who completed the questionnaire believed that playing AquaSnap contributed to their addiction treatment, and 30% (10/34) stated that they would continue playing the game after finishing the treatment. The app's ratings positively correlated with the number of minutes played (Table 5).

**Table 5.** Spearman  $\rho$  correlations between the number of playing minutes with AquaSnap and AquaSnap ratings (n=34).

AquaSnap questions <sup>a</sup>	1	2	3	4	5	6	7
$\rho$	0.8	0.5	0.5	0.5	0.5	0.5	-0.02
<i>P</i> value (2-tailed)	<.001	.002	.001	.008	.005	.004	.93

<sup>a</sup>See Table 4.

### Explorative Analyses Effectiveness AquaSnap

The MyCQ assessment scores and AquaSnap playing time were inspected for outliers by using box plots. In total, 4 MyCQ data points were missing for the cognitive domain *attention*, and 2 were missing for *processing speed*, most likely because of technical failure.

The number of minutes played with AquaSnap correlated with the change in working memory latency time between the first

and second MyCQ assessments; more playing minutes were associated with a decrease in working memory reaction time (Spearman  $\rho=0.4$ ;  $P=.01$ ). An increase in AquaSnap playing minutes was associated with a decrease in executive functioning accuracy between the first and second MyCQ assessments (Spearman  $\rho=-0.3$ ;  $P=.02$ ). No other significant correlations were observed between the first and second assessments (Table 6).

**Table 6.** Spearman  $\rho$  correlations between the number of playing minutes with AquaSnap and change scores<sup>a</sup> between the first (T1) and second (T2) MyCognition Quotient (MyCQ) assessments.

	Attention latency	Attention accuracy	Processing speed latency	Processing speed accuracy	Working memory latency	Working memory accuracy	Episodic memory latency	Episodic memory accuracy	Executive functioning latency	Executive functioning accuracy
<i>n</i>	42	42	44	44	46	46	46	46	46	46
$\rho$	0.1	0.1	0.2	-0.1	0.4 <sup>b</sup>	0.0	0.0	0.2	-0.2	-0.3 <sup>b</sup>
<i>P</i> value	.59	.66	.25	.72	.01 <sup>b</sup>	.88	.84	.22	.14	.02 <sup>b</sup>

<sup>a</sup>Latency (speed in milliseconds) change scores T1 minus T2; accuracy (% true) change scores T2 minus T1.

<sup>b</sup>Correlation is significant at the .03 level (2-tailed).

## Discussion

### Principal Findings

This study investigated the feasibility and acceptability of 2 serious gaming smartphone apps in a group of patients with a SUD in an addiction treatment setting. Both the MyCQ assessment and AquaSnap cognitive training apps were offered in addition to regular addiction care. This study showed that the use of MyCQ is feasible among patients with SUDs who are interested and highly educated. Approximately half of the interested patients actually used MyCQ. However, the feasibility of using the AquaSnap training app was lower with only 25.8% (59/229) of the initially interested patients. This percentage is lower than that in a recent study of patients with SUD on the feasibility of a smartphone attention bias app. Zhang et al [35] investigated the use of an attention bias modification app on a smartphone in 40 inpatients with SUD (predominantly opioid use disorder) and found 75% acceptance and 63% adherence rates. It is important to note that the study by Zhang et al [35] was conducted in a group of inpatients admitted for rehabilitation and lasted only a week. This is in contrast to this

study, in which both inpatients and outpatients were included, who were followed up for 6 weeks. Moreover, participation in this study was entirely voluntary, and participants could stop at any time; therefore, adherence is expected to be higher when implemented as a regular part of treatment.

Highly educated patients were more likely to start using MyCQ. This finding can be partly explained by the fact that highly educated patients are more familiar with scientific research [38]. In addition, highly educated people have better executive functions [39], suggesting that they are better at organizing, staying focused, and exerting self-control [40]. These characteristics promote participation in voluntary studies such as this one. More women than men played the AquaSnap training app. This may indicate volunteer bias [41], although we found no such bias in initial participation in the MyCQ assessment.

Regarding acceptability, our study showed mixed results. The MyCQ assessment can be defined as acceptable for the group that was presented with the questionnaire, with 76% (35/46) and 83% (38/46) of the participants reporting it to be easy and

likable, respectively. Although the AquaSnap game was rated as easy by 68% (23/34) of the participants, only 41% (14/34) enjoyed playing the game. A large majority (34/46, 74%) believed that MyCQ provided insight into their brain functions, whereas only half (16/34, 47%) of the participants believed that playing AquaSnap improved their brain functions. Only 15% (5/34) believed that playing the AquaSnap helped them manage their addiction, which is much lower than the 36% found in the study by Zhang et al [35]. One reason may be that the game used in this study (clicking photographs of fishes) was perceived as less relevant to the addiction problem than the game used in a study by Zhang et al [35] (pushing away substance-related pictures).

The AquaSnap app rating was positively correlated with the number of minutes played (Spearman  $\rho$  ranging from 0.5 to 0.7;  $P < .05$ ). It is possible that the participants played more because they liked the game better, but it could also be that they liked the game better because they played more.

Explorative analysis showed that more playing minutes on AquaSnap was related to shorter reaction times in the domain of working memory but more errors in the domain of executive functioning, as measured by the MyCQ assessment. These associations could be due to factors other than playing the game, such as changes in substance use or the length of substance abstinence. Moreover, the divergent validity of working memory and executive functioning tasks in MyCQ was found to be limited [25]. Therefore, these results should be interpreted with caution.

### Strengths and Limitations

A strength of this study is that it had a naturalistic design and was conducted within regular addiction care, with the feasibility of using a smartphone app investigated in both outpatient and inpatient addiction settings. Patients can play serious games in their own environments at times of their choice. Thus, the ecological validity of this study was high. Participation was low threshold, it took little effort to use the apps, and there were no use restrictions, which increased the number of participants and thus the reliability of the data. Furthermore, this study evaluated the opinions of patients on 2 different smartphone apps, both pertaining to cognitive assessment and cognitive training, providing relevant information for clinical practice.

This study has some limitations. The disadvantage of using a naturalistic design is that there is no standardized implementation of the apps. The circumstances under which the apps were used may have differed, because participants could decide when and where they wanted to use the apps. Furthermore, in addition to the advantages of accessibility, the use of personal phones has several potential complications, such as differences in screen size and processing capacity and speed.

In this study, we defined the recruited population as patients interested in participating. However, we do not know exactly how much of the total population was informed about the study. We know that 41.9% (78/186) of the informed inpatients have participated, but data on this are lacking for the outpatient population. Nevertheless, the participants were comparable with the total population in this department in the year of recruitment

in terms of age, primary SUD, gender, mood, and quality of life. In addition, there was high variability between the participants in terms of the time they spent playing AquaSnap. This hampered the interpretation of the results. Moreover, because participation was voluntary and not an integral part of usual care, adherence rates were low. It is already difficult to monitor addictive patients, because the overall no-shows and dropout rates are known to be high [7]. Another potential confounding factor that we did not measure was a potential difference in familiarity with mobile technology [42]. This may have induced a higher selection of participants with better technological skills, which may have resulted in a more positive evaluation of the apps due to the overinclusion of participants with higher technology readiness.

### Recommendations for Future Research

Future research should include and evaluate the use of smartphone apps as an integral part of usual care. Thus, adherence can be monitored more closely, and there is a greater chance that more people with a lower education level will participate. In addition, patients should be encouraged to train more intensively. More intensive cognitive training has shown positive effects on working memory and alcohol consumption [43,44]. Although the acceptability of the MyCQ assessment was good, the acceptability of the training app AquaSnap was lower. For future research, this app should be made more attractive, or an alternative serious game should be chosen that is also more in line with the addiction problem in terms of content. One of our findings was that the group of patients who used the apps had a higher education level than the group that did not start using the apps. Thus, for feasibility in future studies, attention needs to be paid to engaging patients with different educational levels. Future research using, for example, focus groups could investigate the needs and interests of patients with SUD from different educational levels for these cognitive training interventions.

In addition, it is recommended that initial work be conducted to understand the characteristics of the specific target population and their ownership of smartphone devices [42] before implementation. In clinical practice, the use of the MyCQ app has potential, given the high prevalence of cognitive impairment in this group and the advice to screen for cognitive impairment early in treatment [11]. The fact that app assessment takes an assessor less time than the classic paper-and-pencil test may result in lower costs for cognitive assessment.

### Conclusions

In conclusion, our study shows that the use of a smartphone app for cognitive assessment in patients with SUDs who are interested and highly educated is feasible, and for the subgroup who filled out the questionnaire, it was acceptable. However, our data also highlight that the use of a smartphone app for cognitive training via serious gaming is less feasible in this group of patients. Improvement of the app and motivation of clients to increase the use of serious games is needed. Despite these limitations, the present results provide support for future research investigating the use of smartphone apps for cognitive assessment and cognitive training in relation to the treatment

of SUD because participation and acceptability rates were sufficient.

## Acknowledgments

This research was funded by a *Vidi* grant (91713354) from the Dutch Scientific Organization (Nederlandse Organisatie voor Wetenschappelijk Onderzoek). The authors acknowledge the contributions of the addiction care therapists, Klaas van der Spek and Noa Wielage in recruiting participants and coordinating for the project.

## Authors' Contributions

AEG conceived the study. ESR, MvH, and NG were involved in data collection. ESR and NG contributed to the data entry and coding of the initial data. TS analyzed the data. TS wrote the manuscript with contributions from ESR, MvH, NG, and AEG. All the authors have read and approved the manuscript before submission.

## Conflicts of Interest

None declared.

## References

1. Dutra L, Stathopoulou G, Basden SL, Leyro TM, Powers MB, Otto MW. A meta-analytic review of psychosocial interventions for substance use disorders. *Am J Psychiatry* 2008 Feb;165(2):179-187. [doi: [10.1176/appi.ajp.2007.06111851](https://doi.org/10.1176/appi.ajp.2007.06111851)] [Medline: [18198270](https://pubmed.ncbi.nlm.nih.gov/18198270/)]
2. Jhanjee S. Evidence based psychosocial interventions in substance use. *Indian J Psychol Med* 2014 Apr;36(2):112-118 [FREE Full text] [doi: [10.4103/0253-7176.130960](https://doi.org/10.4103/0253-7176.130960)] [Medline: [24860208](https://pubmed.ncbi.nlm.nih.gov/24860208/)]
3. Heyman GM. Quitting drugs: quantitative and qualitative features. *Annu Rev Clin Psychol* 2013;9:29-59. [doi: [10.1146/annurev-clinpsy-032511-143041](https://doi.org/10.1146/annurev-clinpsy-032511-143041)] [Medline: [23330937](https://pubmed.ncbi.nlm.nih.gov/23330937/)]
4. Cutler RB, Fishbain DA. Are alcoholism treatments effective? The Project MATCH data. *BMC Public Health* 2005 Jul 14;5:75 [FREE Full text] [doi: [10.1186/1471-2458-5-75](https://doi.org/10.1186/1471-2458-5-75)] [Medline: [16018798](https://pubmed.ncbi.nlm.nih.gov/16018798/)]
5. Ciraulo DA, Piechniczek-Buczek J, Iscan E. Outcome predictors in substance use disorders. *Psychiatr Clin North Am* 2003 Jun;26(2):381-409. [doi: [10.1016/s0193-953x\(02\)00106-5](https://doi.org/10.1016/s0193-953x(02)00106-5)]
6. Lappan SN, Brown AW, Hendricks PS. Dropout rates of in-person psychosocial substance use disorder treatments: a systematic review and meta-analysis. *Addiction* 2020 Feb;115(2):201-217. [doi: [10.1111/add.14793](https://doi.org/10.1111/add.14793)] [Medline: [31454123](https://pubmed.ncbi.nlm.nih.gov/31454123/)]
7. Brorson HH, Ajo Arnevik E, Rand-Hendriksen K, Duckert F. Drop-out from addiction treatment: a systematic review of risk factors. *Clin Psychol Rev* 2013 Dec;33(8):1010-1024. [doi: [10.1016/j.cpr.2013.07.007](https://doi.org/10.1016/j.cpr.2013.07.007)] [Medline: [24029221](https://pubmed.ncbi.nlm.nih.gov/24029221/)]
8. Domínguez-Salas S, Díaz-Batanero C, Lozano-Rojas OM, Verdejo-García A. Impact of general cognition and executive function deficits on addiction treatment outcomes: systematic review and discussion of neurocognitive pathways. *Neurosci Biobehav Rev* 2016 Dec;71:772-801. [doi: [10.1016/j.neubiorev.2016.09.030](https://doi.org/10.1016/j.neubiorev.2016.09.030)] [Medline: [27793597](https://pubmed.ncbi.nlm.nih.gov/27793597/)]
9. Sampedro-Piquero P, Ladrón de Guevara-Miranda D, Pavón FJ, Serrano A, Suárez J, Rodríguez de Fonseca F, et al. Neuroplastic and cognitive impairment in substance use disorders: a therapeutic potential of cognitive stimulation. *Neurosci Biobehav Rev* 2019 Nov;106:23-48. [doi: [10.1016/j.neubiorev.2018.11.015](https://doi.org/10.1016/j.neubiorev.2018.11.015)] [Medline: [30481530](https://pubmed.ncbi.nlm.nih.gov/30481530/)]
10. Rolland B, D'Hondt F, Montègue S, Brion M, Peyron E, D'Aviau de Ternay J, et al. A patient-tailored evidence-based approach for developing early neuropsychological training programs in addiction settings. *Neuropsychol Rev* 2019 Mar;29(1):103-115. [doi: [10.1007/s11065-018-9395-3](https://doi.org/10.1007/s11065-018-9395-3)] [Medline: [30607658](https://pubmed.ncbi.nlm.nih.gov/30607658/)]
11. Bruijnen CJ, Dijkstra BA, Walvoort SJ, Markus W, VanDerNagel JE, Kessels RP, et al. Prevalence of cognitive impairment in patients with substance use disorder. *Drug Alcohol Rev* 2019 May;38(4):435-442 [FREE Full text] [doi: [10.1111/dar.12922](https://doi.org/10.1111/dar.12922)] [Medline: [30916448](https://pubmed.ncbi.nlm.nih.gov/30916448/)]
12. Fernández-Serrano MJ, Pérez-García M, Schmidt Río-Valle J, Verdejo-García A. Neuropsychological consequences of alcohol and drug abuse on different components of executive functions. *J Psychopharmacol* 2010 Sep;24(9):1317-1332. [doi: [10.1177/0269881109349841](https://doi.org/10.1177/0269881109349841)] [Medline: [20007413](https://pubmed.ncbi.nlm.nih.gov/20007413/)]
13. Shenassa ED, Paradis AD, Dolan SL, Wilhelm CS, Buka SL. Childhood impulsive behavior and problem gambling by adulthood: a 30-year prospective community-based study. *Addiction* 2012 Jan;107(1):160-168 [FREE Full text] [doi: [10.1111/j.1360-0443.2011.03571.x](https://doi.org/10.1111/j.1360-0443.2011.03571.x)] [Medline: [21752146](https://pubmed.ncbi.nlm.nih.gov/21752146/)]
14. Tarter RE, Kirisci L, Mezzich A, Cornelius JR, Pajer K, Vanyukov M, et al. Neurobehavioral disinhibition in childhood predicts early age at onset of substance use disorder. *Am J Psychiatry* 2003 Jun;160(6):1078-1085. [doi: [10.1176/appi.ajp.160.6.1078](https://doi.org/10.1176/appi.ajp.160.6.1078)] [Medline: [12777265](https://pubmed.ncbi.nlm.nih.gov/12777265/)]
15. Fernández-Serrano MJ, Pérez-García M, Verdejo-García A. What are the specific vs. generalized effects of drugs of abuse on neuropsychological performance? *Neurosci Biobehav Rev* 2011 Jan;35(3):377-406. [doi: [10.1016/j.neubiorev.2010.04.008](https://doi.org/10.1016/j.neubiorev.2010.04.008)] [Medline: [20451551](https://pubmed.ncbi.nlm.nih.gov/20451551/)]

16. van Holst RJ, Schilt T. Drug-related decrease in neuropsychological functions of abstinent drug users. *Curr Drug Abuse Rev* 2011 Mar;4(1):42-56. [doi: [10.2174/1874473711104010042](https://doi.org/10.2174/1874473711104010042)] [Medline: [21466500](https://pubmed.ncbi.nlm.nih.gov/21466500/)]
17. Perry CJ, Lawrence AJ. Addiction, cognitive decline and therapy: seeking ways to escape a vicious cycle. *Genes Brain Behav* 2017 Jan;16(1):205-218 [FREE Full text] [doi: [10.1111/gbb.12325](https://doi.org/10.1111/gbb.12325)] [Medline: [27573875](https://pubmed.ncbi.nlm.nih.gov/27573875/)]
18. Verdejo-Garcia A. Cognitive training for substance use disorders: neuroscientific mechanisms. *Neurosci Biobehav Rev* 2016 Sep;68:270-281. [doi: [10.1016/j.neubiorev.2016.05.018](https://doi.org/10.1016/j.neubiorev.2016.05.018)] [Medline: [27236041](https://pubmed.ncbi.nlm.nih.gov/27236041/)]
19. Lee RS, Hermens DF, Redoblado-Hodge MA, Naismith SL, Porter MA, Kaur M, et al. Neuropsychological and socio-occupational functioning in young psychiatric outpatients: a longitudinal investigation. *PLoS One* 2013;8(3):e58176 [FREE Full text] [doi: [10.1371/journal.pone.0058176](https://doi.org/10.1371/journal.pone.0058176)] [Medline: [23469268](https://pubmed.ncbi.nlm.nih.gov/23469268/)]
20. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. *Mhealth* 2018;4:6 [FREE Full text] [doi: [10.21037/mhealth.2018.03.02](https://doi.org/10.21037/mhealth.2018.03.02)] [Medline: [29682510](https://pubmed.ncbi.nlm.nih.gov/29682510/)]
21. Boendermaker WJ, Veltkamp RC, Peeters M. Training behavioral control in adolescents using a serious game. *Games Health J* 2017 Dec;6(6):351-357. [doi: [10.1089/g4h.2017.0071](https://doi.org/10.1089/g4h.2017.0071)] [Medline: [28926286](https://pubmed.ncbi.nlm.nih.gov/28926286/)]
22. Gamito P, Oliveira J, Lopes P, Brito R, Morais D, Silva D, et al. Executive functioning in alcoholics following an mHealth cognitive stimulation program: randomized controlled trial. *J Med Internet Res* 2014 Apr 17;16(4):e102 [FREE Full text] [doi: [10.2196/jmir.2923](https://doi.org/10.2196/jmir.2923)] [Medline: [24742381](https://pubmed.ncbi.nlm.nih.gov/24742381/)]
23. Verduin ML, LaRowe SD, Myrick H, Cannon-Bowers J, Bowers C. Computer simulation games as an adjunct for treatment in male veterans with alcohol use disorder. *J Subst Abuse Treat* 2013 Mar;44(3):316-322. [doi: [10.1016/j.jsat.2012.08.006](https://doi.org/10.1016/j.jsat.2012.08.006)] [Medline: [23009800](https://pubmed.ncbi.nlm.nih.gov/23009800/)]
24. Gamito P, Oliveira J, Lopes P, Brito R, Morais D, Caçoete C, et al. Cognitive training through mHealth for individuals with substance use disorder. *Methods Inf Med* 2017 Mar 23;56(2):156-161. [doi: [10.3414/ME16-02-0012](https://doi.org/10.3414/ME16-02-0012)] [Medline: [28116414](https://pubmed.ncbi.nlm.nih.gov/28116414/)]
25. Domen AC, van de Weijer SC, Jaspers MW, Denys D, Nieman DH. The validation of a new online cognitive assessment tool: the MyCognition quotient. *Int J Methods Psychiatr Res* 2019 Sep;28(3):e1775 [FREE Full text] [doi: [10.1002/mpr.1775](https://doi.org/10.1002/mpr.1775)] [Medline: [30761648](https://pubmed.ncbi.nlm.nih.gov/30761648/)]
26. Bellens A, Roelant E, Sabbe B, Peeters M, van Dam PA. A video-game based cognitive training for breast cancer survivors with cognitive impairment: a prospective randomized pilot trial. *Breast* 2020 Oct;53:23-32 [FREE Full text] [doi: [10.1016/j.breast.2020.06.003](https://doi.org/10.1016/j.breast.2020.06.003)] [Medline: [32554133](https://pubmed.ncbi.nlm.nih.gov/32554133/)]
27. Guidance on applying for feasibility studies. National Institute for Health Research. URL: <https://www.nihr.ac.uk/documents/nihr-research-for-patient-benefit-rfpb-programme-guidance-on-applying-for-feasibility-studies/20474?pr=> [accessed 2020-12-08]
28. Schippers GM, Broekman TG, Buchholz A, Koeter MW, van den Brink W. Measurements in the Addictions for Triage and Evaluation (MATE): an instrument based on the World Health Organization family of international classifications. *Addiction* 2010 May;105(5):862-871. [doi: [10.1111/j.1360-0443.2009.02889.x](https://doi.org/10.1111/j.1360-0443.2009.02889.x)] [Medline: [20331557](https://pubmed.ncbi.nlm.nih.gov/20331557/)]
29. Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychol Assess* 1998 Jun;10(2):176-181. [doi: [10.1037/1040-3590.10.2.176](https://doi.org/10.1037/1040-3590.10.2.176)]
30. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. *Br J Clin Psychol* 2005 Jun;44(Pt 2):227-239. [doi: [10.1348/014466505X29657](https://doi.org/10.1348/014466505X29657)] [Medline: [16004657](https://pubmed.ncbi.nlm.nih.gov/16004657/)]
31. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Sydney, Australia: Psychology Foundation of Australia; 1995.
32. Priebe S, Huxley P, Knight S, Evans S. Application and results of the Manchester Short Assessment of Quality of Life (MANSA). *Int J Soc Psychiatry* 1999;45(1):7-12. [doi: [10.1177/002076409904500102](https://doi.org/10.1177/002076409904500102)] [Medline: [10443245](https://pubmed.ncbi.nlm.nih.gov/10443245/)]
33. MyCognition. Welldoing. URL: <https://welldoing.org/mycognition> [accessed 2022-08-09]
34. Zhang M, Ying J, Amron SB, Mahreen Z, Song G, Fung DS, et al. A smartphone attention bias intervention for individuals with addictive disorders: protocol for a feasibility study. *JMIR Res Protoc* 2018 Nov 19;7(11):e11822. [doi: [10.2196/11822](https://doi.org/10.2196/11822)]
35. Zhang M, Ying J, Amron SB, Mahreen Z, Song G, Fung DS, et al. A smartphone attention bias app for individuals with addictive disorders: feasibility and acceptability study. *JMIR Mhealth Uhealth* 2019 Sep 12;7(9):e15465 [FREE Full text] [doi: [10.2196/15465](https://doi.org/10.2196/15465)] [Medline: [31586359](https://pubmed.ncbi.nlm.nih.gov/31586359/)]
36. Nouchi R, Taki Y, Takeuchi H, Hashizume H, Akitsuki Y, Shigemune Y, et al. Brain training game improves executive functions and processing speed in the elderly: a randomized controlled trial. *PLoS One* 2012;7(1):e29676 [FREE Full text] [doi: [10.1371/journal.pone.0029676](https://doi.org/10.1371/journal.pone.0029676)] [Medline: [22253758](https://pubmed.ncbi.nlm.nih.gov/22253758/)]
37. van de Weijer SC, Duits AA, Bloem BR, Kessels RP, Jansen JF, Köhler S, et al. The Parkin'Play study: protocol of a phase II randomized controlled trial to assess the effects of a health game on cognition in Parkinson's disease. *BMC Neurol* 2016 Nov 03;16(1):209 [FREE Full text] [doi: [10.1186/s12883-016-0731-z](https://doi.org/10.1186/s12883-016-0731-z)] [Medline: [27809791](https://pubmed.ncbi.nlm.nih.gov/27809791/)]
38. Bauer MW, Shukla R, Allum N. The Culture of Science: How the Public Relates to Science Across the Globe. Milton Park, UK: Routledge; 2012.

39. Huizinga M, Baeyens D, Burack JA. Editorial: executive function and education. *Front Psychol* 2018 Aug 3;9:1357 [FREE Full text] [doi: [10.3389/fpsyg.2018.01357](https://doi.org/10.3389/fpsyg.2018.01357)] [Medline: [30123163](https://pubmed.ncbi.nlm.nih.gov/30123163/)]
40. Diamond A. Executive functions. *Annu Rev Psychol* 2013;64:135-168 [FREE Full text] [doi: [10.1146/annurev-psych-113011-143750](https://doi.org/10.1146/annurev-psych-113011-143750)] [Medline: [23020641](https://pubmed.ncbi.nlm.nih.gov/23020641/)]
41. Rosenthal R, Rosnow RL, Kazdin AE. *Artifacts in Behavioral Research*. Oxford, UK: Oxford University Press; 2009.
42. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. *Adm Policy Ment Health* 2015 Mar;42(2):157-167 [FREE Full text] [doi: [10.1007/s10488-014-0556-2](https://doi.org/10.1007/s10488-014-0556-2)] [Medline: [24824311](https://pubmed.ncbi.nlm.nih.gov/24824311/)]
43. Houben K, Wiers RW, Jansen A. Getting a grip on drinking behavior: training working memory to reduce alcohol abuse. *Psychol Sci* 2011 Jul;22(7):968-975. [doi: [10.1177/0956797611412392](https://doi.org/10.1177/0956797611412392)] [Medline: [21685380](https://pubmed.ncbi.nlm.nih.gov/21685380/)]
44. Rupp CI, Kemmler G, Kurz M, Hinterhuber H, Fleischhacker WW. Cognitive remediation therapy during treatment for alcohol dependence. *J Stud Alcohol Drugs* 2012 Jul;73(4):625-634. [doi: [10.15288/jsad.2012.73.625](https://doi.org/10.15288/jsad.2012.73.625)] [Medline: [22630801](https://pubmed.ncbi.nlm.nih.gov/22630801/)]

## Abbreviations

**DASS:** Depression Anxiety Stress Scale

**MANSA:** Manchester Short Assessment of Quality of Life

**MyCQ:** MyCognition Quotient

**SUD:** substance use disorder

*Edited by A Mavragani; submitted 08.10.21; peer-reviewed by J VanDerNagel, E Caballeria; comments to author 18.12.21; revised version received 10.05.22; accepted 25.05.22; published 06.09.22.*

*Please cite as:*

*Schilt T, Ruijter ES, Godeschalk N, van Haaster M, Goudriaan AE*

*The Use of Smartphone Serious Gaming Apps in the Treatment of Substance Use Disorders: Observational Study on Feasibility and Acceptability*

*JMIR Form Res* 2022;6(9):e34159

URL: <https://formative.jmir.org/2022/9/e34159>

doi: [10.2196/34159](https://doi.org/10.2196/34159)

PMID: [36066923](https://pubmed.ncbi.nlm.nih.gov/36066923/)

©Thelma Schilt, Elvira Sharine Ruijter, Nikky Godeschalk, Marit van Haaster, Anna E Goudriaan. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 06.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Transdiagnostic Internet-Delivered Cognitive Behavioral Therapy for Symptoms of Postpartum Anxiety and Depression: Feasibility Randomized Controlled Trial

Victoria Suchan<sup>1</sup>, PhD; Vanessa Peynenburg<sup>1</sup>, MA; David Thiessen<sup>2</sup>, BSc; Marcie Nugent<sup>1</sup>, MSW; Blake Dear<sup>3</sup>, PhD; Nickolai Titov<sup>3</sup>, PhD; Heather Hadjistavropoulos<sup>1</sup>, PhD

<sup>1</sup>Department of Psychology, University of Regina, Regina, SK, Canada

<sup>2</sup>Department of Mathematics and Statistics, University of Regina, Regina, SK, Canada

<sup>3</sup>Centre Clinic, Department of Psychology, Macquarie University, Sydney, Australia

**Corresponding Author:**

Heather Hadjistavropoulos, PhD

Department of Psychology

University of Regina

3737 Wascana Parkway

Regina, SK, S4S 0A2

Canada

Phone: 1 306 585 5133

Email: [hadjista@uregina.ca](mailto:hadjista@uregina.ca)

## Abstract

**Background:** Postpartum depression (PPD) and postpartum anxiety (PPA) are often comorbid and are associated with significant personal and economic costs. Fewer than half of the mothers experiencing PPD or PPA symptoms receive face-to-face treatment, suggesting a need for alternative delivery formats such as internet-delivered cognitive behavioral therapy (ICBT).

**Objective:** This pilot study aimed to examine the impact of a therapist-assisted, transdiagnostic ICBT program on symptoms of PPD and PPA, as there is only one previous study on transdiagnostic ICBT with this population, which did not include therapist assistance.

**Methods:** Clients endorsing the symptoms of PPD or PPA (N=63) were randomized to an 8-week transdiagnostic ICBT course (*Wellbeing Course for New Moms*) or to treatment as usual (TAU). Clients completed measures of depression, anxiety, stress, postnatal bonding, and relationship satisfaction, as well as measures of treatment satisfaction and therapeutic alliance, before treatment, after treatment, and at the 1-month follow-up. Outcome measures were also completed at the 6-month follow-up for clients who completed the ICBT course.

**Results:** Both the ICBT and TAU groups experienced statistically significant improvements over time. The ICBT group experienced larger improvements after treatment and at the 1-month follow-up on more measures than the TAU group, with medium between-group Cohen *d* effects on primary outcome measures for anxiety (Cohen *d*=0.65, 95% CI 0.13-1.17), PPD (Cohen *d*=0.52, 95% CI 0.01-1.04), and depression (Cohen *d*=0.56, 95% CI 0.05-1.08), and on secondary outcome measures of overall distress (Cohen *d*=0.69, 95% CI 0.17-1.21), anxiety (Cohen *d*=0.59, 95% CI 0.07-1.11), and stress (Cohen *d*=0.76, 95% CI 0.23-1.28). Time-by-group interactions for proportional reductions between groups over time were only significant after treatment and at the 1-month follow-up for the primary anxiety measure (*P*=.006). This study was underpowered for detecting small or medium effects. Overall, clients perceived the treatment as credible, and 95% (21/22) of the clients were satisfied with the treatment content and therapist support.

**Conclusions:** Findings from this pilot study provide preliminary support for transdiagnostic ICBT in treating PPD and PPA symptoms to improve access to psychological treatments.

**Trial Registration:** ClinicalTrials.gov NCT04012580; <https://clinicaltrials.gov/ct2/show/NCT04012580>

(*JMIR Form Res* 2022;6(9):e37216) doi:[10.2196/37216](https://doi.org/10.2196/37216)

**KEYWORDS**

postpartum depression; postpartum anxiety; internet-delivered cognitive behavioral therapy; transdiagnostic; therapist assistance



## Introduction

### Background

In the period following childbirth, women are at an increased risk of experiencing depression and anxiety [1], with up to 20% of women experiencing clinical levels of symptoms [2]. To date, the vast majority of postpartum mental health research has focused on postpartum depression (PPD), leaving postpartum anxiety (PPA) poorly understood and underresearched, despite the high levels of comorbidity between these concerns [3]. The detrimental impacts associated with untreated PPD and PPA, such as impairments in mother-infant bonding [4] and infant development [5] and high economic costs [6], underscore the importance of efficacious treatment options. Cognitive behavioral therapy (CBT) has the largest evidence base for treating symptoms of PPD and PPA [7], and historically, researchers have taken a disorder-specific approach to treatment wherein they focus on a singular disorder in their trials (eg, PPD). Unfortunately, even with effective treatment options available, less than half of the new mothers experiencing symptoms seek treatment [8] because of a myriad of barriers (eg, childcare difficulties, time concerns, stigma, and transportation difficulties [9]). To address the gap between those who require and receive treatment, it is worthwhile to consider alternative treatment modalities such as internet-delivered CBT (ICBT).

### ICBT for PPA and PPD Symptoms

In ICBT, clients work through structured web-based psychoeducational and therapeutic materials similar to what would be covered in face-to-face CBT [10]. ICBT is an effective treatment for a range of mood and anxiety disorders, yielding similar effects to face-to-face CBT [11]. The delivery of ICBT can vary across several dimensions, including the type of therapist guidance (self-guided vs therapist-assisted) and specificity of content (disorder-specific vs transdiagnostic). In a meta-analysis of ICBT for nonpostpartum clients, therapist-assisted programs were superior to self-guided programs [12]. Although a growing body of literature exists demonstrating the effectiveness of ICBT in treating symptoms of anxiety [13] and depression [14] in pregnant women, as well as depression in the postpartum period [8,15,16], there is much less research on PPA. There is only one known ICBT study that specifically recruited postpartum women experiencing anxiety [17]; unfortunately, in the treatment group, only 2 treatment participants accessed all modules, and only 17% completed posttreatment questionnaires. Although this study provides helpful information, its limitations suggest that these results should be interpreted with caution.

The other notable gap in the literature is that the research on PPA and PPD symptoms has historically been disorder specific in nature (ie, recruiting women with, and providing interventions specifically for, PPA or PPD); however, the high rates of comorbidity between PPD and PPA symptoms [3] suggest that transdiagnostic ICBT may be particularly beneficial in addressing comorbid concerns efficiently. To date, only 1 study has used a transdiagnostic ICBT framework to address the symptoms of PPD or PPA [18]. The *MUMentum Postnatal*

program was adapted from an existing ICBT program for adults with anxiety and depression. Adaptations included shortening the number of lessons and course duration, modifying psychoeducation throughout to be specifically relevant to postpartum concerns, and providing additional postpartum-relevant resources to accompany each lesson. The program comprised 3 core lessons tailored to new mothers. The program was intended to be completed over a span of 4 weeks and was entirely self-guided, with therapists only contacting new mothers if there were significant elevations in symptoms or suicidality. New mothers (N=87) were randomly assigned to receive either the *MUMentum Postnatal* program or treatment as usual (TAU). After treatment and at follow-up, new mothers assigned to the treatment group experienced larger improvements in the measures of anxiety, depression, and distress than the new mothers who received TAU. Furthermore, there were moderate but significant differences in measures of bonding and quality of life in favor of the *MUMentum Postnatal* program over TAU. Adherence rates were high, with 75% of the new mothers completing all 3 lessons, and new mothers were satisfied with the treatment overall. This trial provides preliminary evidence for the efficacy and acceptability of self-guided transdiagnostic ICBT for PPD and PPA. Additional research examining the inclusion of therapist assistance in transdiagnostic ICBT for PPD and PPA is warranted, given the finding that programs with therapist assistance often facilitate greater symptom reduction than self-guided programs [12].

### Objectives

The aim of this pilot study, which was planned before the abovementioned study was published, was to examine the efficacy of a therapist-assisted transdiagnostic program (ie, the *Wellbeing Course for New Moms*) for new mothers experiencing symptoms of PPD or PPA when delivered by a clinic specializing in the provision of ICBT and providing services on a continuous basis funded by the government. Examining the program in a routine care setting allows for greater generalizability of findings outside the initial research setting in which a program is developed [19,20]. Specifically, this study compared the effects of the *Wellbeing Course for New Moms* with TAU in reducing symptoms of anxiety and depression and improving new mothers' rated stress, bonding with their infants, and relationship satisfaction. We took an approach similar to that of Loughnan et al [18], in that a pre-existing ICBT program was modified for new mothers, in this case, by adding an additional supplementary resource. The *MUMentum Postnatal* and *Wellbeing Course for New Moms* programs differ in terms of length (4 weeks vs 8 weeks), number of lessons (3 lessons vs 5 lessons), therapist support (unguided vs therapist guided), and specific outcome measures (eg, for distress). Another objective of this study was to examine new mothers' satisfaction with the course. The results of this study can also serve as a replication of the findings of the *MUMentum Postnatal* program [18], which is important for establishing the generalizability of ICBT for new mothers in different settings [21]. It was hypothesized that new mothers who received the *Wellbeing Course for New Moms* would show greater improvements in anxiety, depression, stress, bonding with their infants, and relationship satisfaction and that, overall, new mothers would

rate the intervention as acceptable. Given the dearth of research examining transdiagnostic ICBT tailored to new mothers, this pilot study offers important insights into how web-based interventions can help address the mental health needs of new mothers.

## Methods

### Study Context and Design

This pilot study was conducted through the Online Therapy Unit at the University of Regina. The Online Therapy Unit is a publicly funded specialized clinic that offers free ICBT to residents of Saskatchewan experiencing symptoms of anxiety or depression. The clients were randomized to receive the *Wellbeing Course for New Moms* or to TAU.

### Ethics Approval

Ethics approval was obtained from the University of Regina Ethics Board (approval number: 2019-077), and the trial was registered at ClinicalTrials.gov (NCT04012580). Participants provided consent at various points during the study, including completing a consent form before filling out the web-based screening, providing verbal consent during the telephone screening, and before initiating the *Wellbeing Course for New Moms*.

### Participants and Recruitment

Clients were recruited through web-based advertisements, presentations to health care professionals, and printed promotional materials over a span of 6 months (September 2019 to March 2020). Clients began by visiting the Online Therapy Unit website, where they completed a web-based consent form, initial eligibility questions, and a web-based screening questionnaire. To be eligible for this study, potential clients had to (1) be aged  $\geq 18$  years, (2) be female, (3) have given birth and have a child aged  $< 1$  year, (4) have a score  $\geq 10$  on the Edinburgh Postnatal Depression Scale (EPDS; [22]) or score  $\geq 9$  on the 7-item Generalized Anxiety Disorder (GAD-7) questionnaire [23], (5) be a resident of Saskatchewan, (6) be comfortable using technology, (7) have access to a secure computer and the internet, and (8) be willing to provide a medical contact as an emergency contact. Potential clients were excluded if they did not meet the abovementioned inclusion criteria or if they (1) were hospitalized in the prior year for mental health concerns or suicidality; (2) had unmanaged alcohol or drug use, mania, or psychosis; or (3) started a new psychotropic medication within the past month.

Clients who met the initial eligibility criteria completed a telephone intake interview with a doctoral student in clinical psychology (VS) or a master's level social worker (KA). Immediately after being accepted into the trial, clients were randomly assigned to ICBT or TAU. Before recruitment for this study, the primary investigator (VS) generated a random allocation sequence using a randomizer website and specified a 1:1 ratio in blocks of 50. The sequence was produced in the form of a Microsoft Excel file, which was then uploaded directly to the website hosting the web-based screening (REDCap [Research Electronic Data Capture]; Vanderbilt University). After this point, the researchers were unable to view the

allocation sequence. Randomization was performed after participants were deemed eligible for this study to limit the possibility of bias in conducting the telephone screen in the event that the randomization group was known. At the end of the telephone screen, the screener informed participants of their allocation.

### Intervention

The *Wellbeing Course for New Moms* is a transdiagnostic course based on the principles of CBT. It was adapted from the *Wellbeing Course* [24] to reflect the common concerns faced by new mothers. Clients accessed the course through the Online Therapy Unit web platform. The content of any messages and questionnaire responses were encrypted using Advanced Encryption Standard encryption with 256-bit key length. Clients completed 5 web-based lessons spanning over the course of 8 weeks, with weekly support from a therapist. The lessons resemble a Microsoft PowerPoint presentation, which participants can read through independently and move through at their own pace. Lesson 1 (1 week) provides psychoeducation about anxiety and depression in general and in postpartum populations; description of symptoms; and explanation of the relationship between unhelpful thoughts, physical symptoms, and unhelpful behaviors. Lesson 2 (2 weeks) provides information on unhelpful thoughts in relation to the CBT model and strategies for monitoring and challenging thoughts. Lesson 3 (1 week) comprises psychoeducation on physical symptoms in relation to the CBT model and strategies for managing both under- and overarousal (eg, controlled breathing). Lesson 4 (2 weeks) focuses on information related to unhelpful behaviors in relation to the CBT model and guidelines about behavioral activation and graded exposure. The fifth and final lesson (2 weeks) includes information about relapse prevention, normalization, and creation of relapse prevention plans. Each lesson includes case stories and do-it-yourself guides that were modified to be relevant to new mothers to promote the practice of strategies from the course, as well as additional resources that could be accessed at any point throughout the course (ie, assertiveness, managing beliefs, communication, mental skills, managing panic attacks, managing posttraumatic stress disorder, sleep hygiene, problem-solving and worry time, and balancing new motherhood). A new resource (ie, balancing new motherhood) was created for the purpose of this study to specifically provide information about PPD and PPA, as well as common struggles that new mothers face (eg, limited sleep, new roles, isolation, and low self-esteem). The resource was created by one of the authors (VS) and then revised before use based on feedback from the coauthor (HH), 2 psychologists with young children, and 8 mothers recruited from the community.

### Therapist Support

All clients assigned to ICBT received weekly therapist support from a master-level, registered social worker trained in the provision of ICBT. The therapist contacted clients on the same day each week throughout the 8-week course using secure emails on the Online Therapy Unit's platform. In this way, communication between the client and therapist was asynchronous, meaning that the sender's message may not be

read until several days later by the receiver. Each therapist message was personalized but included several important elements: warmth and concern, feedback on symptom measures, highlighting key skills, answering client questions about the lessons, acknowledging any stated areas of difficulty, providing encouragement reinforcing progress, managing risk, and reminding clients about course instructions (see the study by Hadjistavropoulos et al [25] for more details). Phone calls were made to clients in specific cases (ie, heightened risk of suicide, significant increase in symptoms of anxiety or depression, or to clarify client concerns). The clients also received automated messages as reminders of new lessons or questionnaires to complete. Clients received an average of 9.89 (SD 1.03) messages and 1.93 (SD 1.96) phone calls from their therapists and sent an average of 4.32 (SD 5.23) messages.

## Measures

Outcome measures were administered before treatment, after treatment, and at the 1-month follow-up. Clients who were assigned to ICBT also completed the outcome measures at the 6-month follow-up. The measures took approximately 15 minutes to complete.

### Primary Measures

#### EPDS Questionnaire

The EPDS [22] is a psychometrically sound 10-item self-report questionnaire that assesses symptoms of depression and anxiety in new mothers. Total scores on the EPDS range from 0 to 30, with higher scores representing more severe symptoms. Consistent with previous research [17], a cutoff of  $\geq 10$  was considered suggestive of clinical levels of depressive symptoms. The  $\alpha$  range for the EPDS in the current trial was .86 to .97.

#### GAD-7 Measure

The GAD-7 [23] is a psychometrically sound 7-item measure that assesses symptoms of generalized anxiety and has been validated in perinatal samples [26]. Total scores range from 0 to 21, with higher scores indicating more severe symptoms of generalized anxiety. A cutoff score of 9 was used, consistent with previous trials of ICBT for PPD and PPA [18]. The  $\alpha$  range for the GAD-7 in this trial was .75 to .92.

#### 9-Item Patient Health Questionnaire

The 9-item Patient Health Questionnaire (PHQ-9; [27]) is a psychometrically sound 9-item measure that assesses depressive symptoms and has been validated in postpartum populations [28]. Total scores range from 0 to 27, with higher scores indicating more severe depressive symptoms. The  $\alpha$  range for the PHQ-9 in this trial was .67 to .90.

### Secondary Measures

#### Depression and Anxiety Stress Scale

The 21-item Depression and Anxiety Stress Scale (DASS-21; [29]) is a psychometrically sound measure that comprises 21 items assessing 3 subscales: depression, stress, and anxiety. The range for total scores on the DASS-21 is 0 to 63, and the range for each subscale is 0 to 21, with higher scores indicating more severe symptoms. The  $\alpha$  range for the DASS-21 in this trial was .89 to .94.

#### Postnatal Bonding Questionnaire

The Postnatal Bonding Questionnaire (PBQ; [30]) is a psychometrically sound 25-item self-report questionnaire that assesses mothers' difficulties with bonding, rejection and anger, anxiety about care, and the risk of abuse regarding their infant. Total scores range from 0 to 125, with higher scores indicating more difficulties with postnatal bonding. The  $\alpha$  range for the PBQ in this trial was .72 to .91.

#### 7-Item Dyadic Adjustment Scale

The 7-item Dyadic Adjustment Scale (DAS-7; [31]) is a psychometrically sound 7-item measure that assesses perceived relationship quality. Total scores on the DAS-7 range from 0 to 36, with higher scores indicating greater perceived relationship quality. The  $\alpha$  range for DAS-7 in this trial was .73 to .91.

#### Treatment Satisfaction, Credibility, and Working Alliance

##### Treatment Satisfaction Questionnaire

The Treatment Satisfaction Questionnaire (TSQ; [32]) was created by the eCentre Clinic and has been used in several ICBT studies [32]. Clients completed the TSQ after treatment regarding their satisfaction with the *Wellbeing Course for New Moms*. The TSQ comprises 6 questions to assess satisfaction with the treatment (*very dissatisfied* to *very satisfied*), satisfaction with the quality of the lessons and do-it-yourself guides (*very dissatisfied* to *very satisfied*), confidence in recommending the treatment to a friend (*yes* or *no*), whether the treatment was worth their time (*yes* or *no*), how participating in treatment affected their confidence to manage symptoms (*greatly reduced* to *greatly increased*), and how participating in treatment affected their motivation to seek future treatment if needed (*greatly reduced* to *greatly increased*).

##### Adverse Effects Questionnaire

After the treatment, clients were asked to indicate whether they experienced any negative effects, new psychological symptoms, or unwanted events (eg, family member death or loss of a job) during the course. If clients answered *yes* to these questions, they were asked to explain further. Clients were also asked to indicate whether that event negatively affected their participation, engagement, or benefit from the program.

##### Credibility and Expectancy Questionnaire

The Credibility and Expectancy Questionnaire (CEQ; [33]) comprises 6 items to assess beliefs about treatment credibility and expectancy. It was administered before treatment to both groups and after treatment to the ICBT treatment group. Total scores for each of the 2 factors (ie, credibility and expectancy) ranged from 3 to 27. The  $\alpha$  range was .60 to .88 for the credibility scale and .70 to .84 for the expectancy scale of the CEQ.

##### Working Alliance Inventory-Short Revised

The Working Alliance Inventory-Short Revised [34] was administered to the ICBT treatment group after treatment. It comprises 12 items to assess the therapeutic relationship and 3 subscales that examine the bond between therapist and client,

agreement on goals, and agreement on tasks in therapy. Items on each of the 3 subscales can be summed to create a total subscale score ranging from 4 to 20. The 3 subscales can be summed to create a total working alliance score ranging from 20 to 60. The  $\alpha$  range for the Working Alliance Inventory-Short Revised in this trial was .72 to .90.

### ***Health Service Use: Service Utilization Questionnaire***

All clients were asked whether they used a number of different services during the previous 8-week period at the posttreatment and 1-month follow-up (ie, medical physician, psychologist or other mental health care worker, support group, psychotropic medications, naturopathic medicine, naturopathic or homeopathic procedures, or other treatments for PPD or PPA) time points. Clients in the ICBT treatment group also received these questions at the 6-month follow-up.

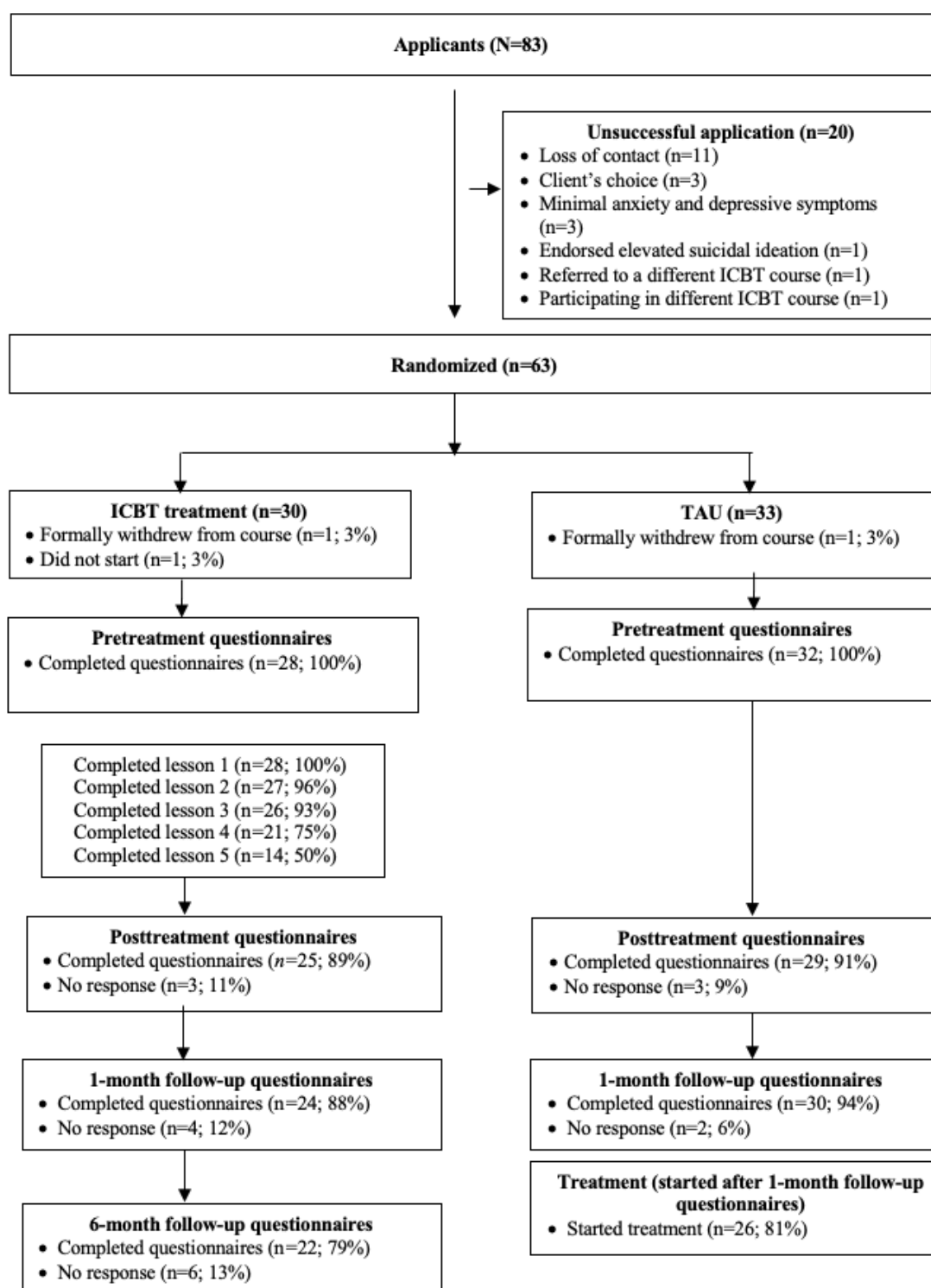
### **Statistical Analysis**

Descriptive statistics were calculated for demographic and clinical characteristics to summarize the sample and check for pretreatment differences between the conditions. Changes in outcome measures over time were modeled using generalized estimating equations (GEEs; [35]). For all GEE models, we used an exchangeable working correlation and robust “sandwich” estimates of SEs. Gamma distributions were used to accommodate the observed skewed responses. We used a log-link function, which leads to regression coefficients and

hypothesis tests assuming that changes are proportional to pretreatment measures. Modeling changes as proportional reductions from pretreatment has been recommended to provide a better model fit and reduce measurement error.

To compare improvements from before treatment to after treatment and at the 1-month follow-up, we calculated proportional improvements from the pretreatment time point and within-group and between-group Cohen *d* effect sizes. Proportional improvements and within-group Cohen *d* effect sizes are also reported for the ICBT treatment group at the 6-month follow-up. Hypothesis tests of differences in improvements between groups were performed using Wald tests on the estimated time $\times$ group interaction coefficients from the GEE models.

Questionnaire completion rates were high in both groups at all observation times (Figure 1). However, ignoring missing cases can lead to an overestimation of treatment effects [36]. Therefore, we used multiple imputation to replace missing outcome measures, controlling for pretreatment symptom severity and course completion rates. In the TAU, course completion was not available; rather, we controlled for pretreatment symptom severity and whether the client accessed other mental health supports (ie, general practitioner, mental health worker, or other support groups), treating the mental health support access as a proxy measure for course engagement.

**Figure 1.** Participant flow. ICBT: internet-delivered cognitive behavioral therapy; TAU: treatment as usual.

## Results

### Demographic and Clinical Characteristics

Over the 6-month recruitment period, 83 individuals applied to the *Wellbeing Course for New Moms*. Of these 83 individuals, 20 (24%) applications were considered unsuccessful for the following reasons: loss of contact (n=11, 55%), client's choice (n=3, 15%), minimal symptoms of anxiety and depression (n=3, 15%), elevated suicidal ideation (n=1, 5%), referred to

web-based materials or resources (n=1, 5%), or participating in another ICBT course (ie, *Wellbeing Course*; n=1, 5%). The remaining 63 clients were randomized to either transdiagnostic ICBT (n=30, 48%) or TAU (n=33, 52%). In the ICBT group, 3% (1/30) of clients did not start the course, and 3% (1/30) of clients formally withdrew. In the TAU group, 3% (1/33) of clients formally withdrew before completing the pretreatment questionnaires. Completion of posttreatment questionnaires (ICBT: 25/28, 89%; TAU: 29/32, 91%), 1-month follow-up

questionnaires (ICBT: 24/28, 86%; TAU: 30/32, 94%), and 6-month follow-up questionnaires (ICBT: 22/28, 79%; TAU: not administered) was high. [Figure 1](#) provides a more detailed description of the client flow.

The average age of the clients was 30.83 (SD 4.29) years. Most clients were White (51/60, 85%), married or in a common law relationship (56/60, 93%), and on maternity or parental leave (45/60, 75%). Furthermore, most clients indicated some level of postsecondary education (50/60, 83%). The average age of the clients' infants was 4.63 (SD 3.61) months, and 60% (36/60) of the clients reported having other children. The most common type of reported birth was vaginal (40/60, 67%), followed by unplanned cesarean (12/60, 20%) or planned cesarean (8/60,

13%) sections. Additional demographic information is presented in [Table 1](#). No statistically significant differences were found in any demographic characteristics between the ICBT treatment and TAU groups ( $P=.13-.99$ ).

At intake, clients' average scores on the EPDS fell within the range of probable depression (mean 14.98, SD 4.54) and within the moderate range for both the GAD-7 (mean 12.47, SD 5.27) and PHQ-9 (mean 11.33, SD 5.17). The mean scores on the PBQ and DAS-7 were 18.03 (SD 11.67) and 23.88 (SD 4.88), respectively. The only significant group difference that emerged was on the DAS-7 ( $t_{56}=2.98$ ;  $P=.01$ ), whereby clients in the TAU group reported greater relationship satisfaction.

**Table 1.** Descriptive statistics for demographic variables (N=60).

Variables	All participants	ICBT <sup>a</sup> (n=28)	TAU <sup>b</sup> (n=32)
Age (years), mean (SD)	30.83 (4.29)	31.18 (4.31)	30.53 (4.32)
Infant's age (months), mean (SD)	4.63 (3.12)	4.96 (2.93)	4.34 (3.13)
<b>Type of birth, n (%)</b>			
Vaginal	40 (67)	20 (71)	20 (63)
Planned cesarean section	8 (13)	4 (14)	4 (13)
Unplanned cesarean section	12 (20)	4 (14)	8 (25)
<b>Obstetric complications, n (%)</b>			
No	39 (65)	17 (61)	22 (69)
Yes	21 (35)	11 (39)	10 (31)
<b>Breastfeeding status, n (%)</b>			
Breastfeeding exclusively	32 (53)	14 (50)	18 (56)
Breastfed initially, now using formula	14 (23)	9 (32)	5 (16)
Breastfeeding and formula	8 (13)	3 (11)	5 (16)
Formula exclusively since birth	6 (10)	2 (7)	4 (13)
<b>Number of children, n (%)</b>			
1	24 (40)	11 (39)	13 (41)
2	21 (35)	6 (21)	15 (47)
3	8 (13)	7 (25)	1 (3)
≥4	7 (12)	4 (14)	3 (9)
<b>Current episode onset, n (%)</b>			
Before becoming pregnant	16 (27)	9 (32)	7 (22)
During pregnancy	15 (25)	8 (29)	7 (22)
Within 1 month after birth	17 (28)	5 (18)	12 (38)
2-4 months after birth	5 (8)	3 (11)	2 (6)
5-7 months after birth	7 (12)	3 (11)	4 (13)
<b>Ethnicity, n (%)</b>			
White	51 (85)	21 (75)	30 (94)
Aboriginal (First Nations, Métis)	5 (8)	5 (18)	0
Other	4 (7)	2 (7)	2 (6)
<b>Employment status, n (%)</b>			
Working	3 (5)	1 (4)	2 (6)
Student	3 (5)	2 (7)	1 (3)
On maternity leave	45 (75)	21 (75)	24 (75)
Not working or on disability	9 (15)	4 (14)	5 (16)
<b>Family's annual income (US \$), n (%)</b>			
10,000-24,999	1 (2)	0	1 (3)
25,000-49,000	7 (12)	4 (14)	3 (9)
50,000-74,000	7 (12)	2 (7)	5 (16)
75,000-99,000	11 (18)	6 (21)	5 (16)
100,000-149,000	20 (33)	9 (32)	11 (34)
≥150,000	10 (17)	6 (21)	4 (13)
Prefer not to disclose or do not know	4 (7)	1 (4)	3 (9)

Variables	All participants	ICBT <sup>a</sup> (n=28)	TAU <sup>b</sup> (n=32)
<b>Education, n (%)</b>			
High school diploma	11 (18)	5 (18)	6 (19)
College or some university	21 (35)	7 (25)	14 (44)
Undergraduate degree	20 (33)	8 (29)	12 (38)
Professional degree (eg, MD)	4 (7)	4 (14)	0
Graduate degree (eg, MA or PhD)	4 (7)	4 (14)	0
<b>Relationship status, n (%)</b>			
Married or common law	56 (93)	26 (93)	30 (94)
In a relationship	2 (3)	1 (4)	1 (3)
Separated or single	2 (3)	1 (4)	1 (3)
<b>Location, n (%)</b>			
Large city (>200,000)	24 (40)	11 (39)	13 (41)
Small city (10,000-200,000)	16 (27)	5 (18)	11 (34)
Town or village or farm	20 (33)	13 (43)	8 (25)
<b>On psychological medication, n (%)</b>			
Yes	24 (40)	13 (46)	11 (34)
No	36 (60)	15 (54)	21 (66)
<b>Service use after treatment<sup>c</sup> (n=51), n (%)</b>			
Medical physician	14 (27)	4 (18)	10 (35)
Psychologist or mental health worker	14 (27)	5 (23)	9 (31)
Support group	9 (18)	5 (23)	4 (14)
Initiation of psychological medication	10 (20)	3 (14)	7 (24)
Naturopathic or homeopathic procedure	4 (8)	3 (14)	1 (3)
<b>Service use at the 1-month follow-up<sup>c</sup> (n=54), n (%)</b>			
Medical physician	9 (17)	3 (13)	6 (20)
Psychologist or mental health worker	17 (32)	8 (33)	9 (30)
Support group	4 (7)	2 (7)	2 (7)
Initiation of psychological medication	9 (17)	4 (17)	5 (17)
Naturopathic or homeopathic procedure	8 (15)	4 (17)	4 (13)
<b>Service use at the 6-month follow-up<sup>c</sup> (n=22), n (%)</b>			
Medical physician	1 (5)	1 (5)	__ <sup>d</sup>
Psychologist or mental health worker	5 (23)	5 (23)	—
Support group	0 (0)	0 (0)	—
Initiation of psychological medication	0 (0)	0 (0)	—
Naturopathic or homeopathic procedure	1 (5)	1 (5)	—

<sup>a</sup>ICBT: internet-delivered cognitive behavioral therapy.

<sup>b</sup>TAU: treatment as usual.

<sup>c</sup>Service use includes any visits for mental health over the previous 8 weeks.

<sup>d</sup>Not available.

## Primary Outcomes

Table 2 includes the means, SDs, and proportional reductions for the EPDS, GAD-7, and PHQ-9 for the groups from before

treatment to after treatment and at the 1- and 6-month follow-ups. GEE analyses revealed that the ICBT treatment group experienced large improvements at all time points on the EPDS (proportional reductions 34%-46%; Cohen  $d=0.98-1.47$ ),



GAD-7 (proportional reductions 51%-63%; Cohen  $d=1.41-2.03$ ), and PHQ-9 (proportional reductions 41%-56%; Cohen  $d=0.89-1.37$ ). The TAU group experienced medium improvements after treatment and at the 1-month follow-up on the EPDS (proportional reductions 20%-25%; Cohen  $d=0.62-0.72$ ), GAD-7 (proportional reductions 18%-30%; Cohen  $d=0.40-0.64$ ), and PHQ-9 (proportional reductions 27%-31%; Cohen  $d=0.66-0.68$ ). As indicated in [Table 3](#), there were medium between-group effect sizes both after treatment and at the 1-month follow-up on all primary measures (Cohen  $d=0.52-0.65$ ), with the ICBT treatment group having better

outcomes. Hypothesis tests on time $\times$ group interactions showed that the differences in proportional improvements were only statistically significant on the GAD-7 ( $P=.006$ ), with the ICBT treatment group having larger improvements. Tests were not significant on the EPDS ( $P=.20$ ) or PHQ-9 ( $P=.16$ ). Clients in the ICBT treatment group experienced further improvements on the EPDS (proportional reductions 32%-56%; Cohen  $d=0.88-2.06$ ), GAD-7 (proportional reductions 53%-71%; Cohen  $d=1.39-2.68$ ), and PHQ-9 (proportional reductions 41%-67%; Cohen  $d=0.79-1.95$ ) at the 6-month follow-up.

**Table 2.** Estimated marginal means, SDs, and percentage changes for primary and secondary outcomes by group.

Outcomes	Estimated marginal means (SD)				Percentage changes from pretreatment questionnaires			
	Pretreatment	Posttreatment	1-month follow-up	6-month follow-up	Posttreatment	1-month follow-up	6-month follow-up (95% CI)	
					95% CI	<i>P</i> value	95% CI	<i>P</i> value
<b>Primary outcomes</b>								
<b>EPDS<sup>a</sup></b>						.13		.06
ICBT <sup>b</sup>	14.47 (4.27)	9.54 (5.59)	8.75 (4.38)	7.86 (4.58)	34 (17 to 48)		40 (27 to 50)	46 (32 to 57)
TAU <sup>c</sup>	15.44 (4.79)	12.35 (5.04)	11.54 (5.82)	— <sup>d</sup>	20 (8 to 31)		25 (11 to 37)	—
<b>GAD-7<sup>e</sup></b>						.002		.006
ICBT	13.36 (5.00)	6.51 (4.56)	5.79 (3.90)	4.97 (2.87)	51 (36 to 63)		57 (43 to 67)	63 (53 to 71)
TAU	11.69 (5.46)	9.58 (4.80)	8.24 (5.26)	—	18 (2 to 31)		30 (12 to 44)	—
<b>PHQ-9<sup>f</sup></b>						.15		.06
ICBT	10.68 (5.11)	6.29 (4.64)	5.53 (3.18)	4.74 (3.23)	41 (22 to 56)		48 (35 to 59)	56 (41 to 67)
TAU	11.91 (5.24)	8.73 (3.94)	8.23 (5.75)	—	27 (14 to 38)		31 (12 to 46)	—
<b>Secondary outcomes</b>								
<b>DASS-21<sup>g</sup> total</b>						.04		.02
ICBT	24.14 (10.44)	13.03 (9.80)	12.17 (7.64)	9.91 (7.94)	46 (26 to 61)		50 (36 to 60)	59 (44 to 70)
TAU	26.25 (12.85)	20.24 (10.73)	18.57 (10.09)	—	23 (7 to 36)		29 (15 to 41)	—
<b>DASS-21 Depression</b>						.32		.11
ICBT	6.83 (5.11)	4.08 (3.59)	3.96 (2.95)	2.87 (3.74)	40 (14 to 59)		42 (23 to 56)	58 (30 to 75)
TAU	7.57 (5.22)	5.62 (4.48)	5.98 (5.24)	—	26 (2 to 44)		21 (–8 to 42)	—
<b>DASS-21 Anxiety</b>						.07		.13
ICBT	6.08 (3.41)	2.61 (3.49)	2.45 (2.87)	1.93 (2.22)	57 (24 to 76)		60 (35 to 75)	68 (50 to 80)
TAU	6.57 (5.17)	4.94 (4.21)	4.00 (3.07)	—	25 (–1 to 44)		39 (21 to 53)	—
<b>DASS-21 Stress</b>						.03		.02
ICBT	11.25 (4.02)	6.61 (3.92)	5.96 (3.26)	5.17 (3.22)	41 (25 to 54)		47 (34 to 57)	54 (42 to 64)
TAU	12.13 (4.79)	9.68 (4.07)	8.75 (4.39)	—	20 (7 to 31)		28 (14 to 39)	—
<b>PBQ<sup>h</sup></b>						.63		.70
ICBT	17.25 (10.38)	12.05 (6.97)	10.58 (4.88)	8.70 (7.43)	30 (11 to 45)		39 (26 to 49)	50 (29 to 64)
TAU	18.72 (12.82)	14.27 (10.03)	12.32 (9.21)	—	24 (2 to 41)		34 (15 to 49)	—
<b>DAS-7<sup>i</sup></b>						.04		.63
ICBT	21.96 (3.72)	23.37 (3.85)	22.81 (4.29)	24.00 (4.78)	10 (2 to 21)		6 (–7 to 18)	15 (0 to 27)
TAU	25.55 (5.21)	25.09 (5.41)	25.76 (5.52)	—	–4 (–24 to 12)		2 (–18 to 19)	—

<sup>a</sup>EPDS: Edinburgh Postnatal Depression Scale.

<sup>b</sup>ICBT: internet-delivered cognitive behavior therapy.

<sup>c</sup>TAU: treatment as usual.

<sup>d</sup>Not available.

<sup>e</sup>GAD-7: 7-item Generalized Anxiety Disorder.

<sup>f</sup>PHQ-9: 9-item Patient Health Questionnaire.

<sup>g</sup>DASS-21: 21-item Depression and Anxiety Stress Scale.

<sup>h</sup>PBQ: Postnatal Bonding Questionnaire.

<sup>i</sup>DAS-7: 7-item Dyadic Adjustment Scale.

**Table 3.** Within-group and between-group effect sizes for primary and secondary outcomes.

Outcomes	Within-group effect sizes from pretreatment questionnaires (95% CI)			Between-group effect sizes from pretreatment questionnaires (95% CI)	
	Posttreatment	1-month follow-up	6-month follow-up	Posttreatment	1-month follow-up
<b>Primary outcomes</b>					
<b>EPDS<sup>a</sup></b>					
ICBT <sup>b</sup>	0.98 (0.42 to 1.53)	1.30 (0.73 to 1.88)	1.47 (0.88 to 2.06)	0.52 (0.01 to 1.04)	0.53 (0.01 to 1.05)
TAU <sup>c</sup>	0.62 (0.12 to 1.12)	0.72 (0.22 to 1.23)	N/A <sup>d</sup>	0.52 (0.01 to 1.04)	0.53 (0.01 to 1.05)
<b>GAD-7<sup>e</sup></b>					
ICBT	1.41 (0.83 to 2.00)	1.67 (1.06 to 2.27)	2.03 (1.39 to 2.68)	0.65 (0.13 to 1.17)	0.52 (0.00 to 1.03)
TAU	0.40 (-0.09 to 0.90)	0.64 (0.13 to 1.14)	N/A	0.65 (0.13 to 1.17)	0.52 (0.00 to 1.03)
<b>PHQ-9<sup>f</sup></b>					
ICBT	0.89 (0.34 to 1.44)	1.19 (0.62 to 1.76)	1.37 (0.79 to 1.95)	0.56 (0.05 to 1.08)	0.56 (0.05 to 1.08)
TAU	0.68 (0.17 to 1.18)	0.66 (0.16 to 1.16)	N/A	0.56 (0.05 to 1.08)	0.56 (0.05 to 1.08)
<b>Secondary outcomes</b>					
<b>DASS-21<sup>g</sup> total</b>					
ICBT	1.08 (0.52 to 1.64)	1.29 (0.72 to 1.87)	1.51 (0.92 to 2.11)	0.69 (0.17 to 1.21)	0.70 (0.18 to 1.22)
TAU	0.50 (0.00 to 1.00)	0.66 (0.15 to 1.16)	N/A	0.69 (0.17 to 1.21)	0.70 (0.18 to 1.22)
<b>DASS-21 Depression</b>					
ICBT	0.62 (0.08 to 1.15)	0.68 (0.14 to 1.22)	0.87 (0.32 to 1.42)	0.37 (-0.14 to 0.88)	0.46 (-0.05 to 0.98)
TAU	0.40 (-0.10 to 0.89)	0.30 (-0.19 to 0.79)	N/A	0.37 (-0.14 to 0.88)	0.46 (-0.05 to 0.98)
<b>DASS-21 Anxiety</b>					
ICBT	0.99 (0.44 to 1.55)	1.14 (0.57 to 1.70)	1.42 (0.84 to 2.01)	0.59 (0.07 to 1.11)	0.51 (0.00 to 1.03)
TAU	0.34 (-0.15 to 0.84)	0.60 (0.10 to 1.10)	N/A	0.59 (0.07 to 1.11)	0.51 (0.00 to 1.03)
<b>DASS-21 Stress</b>					
ICBT	1.15 (0.59 to 1.72)	1.42 (0.84 to 2.01)	1.65 (1.04 to 2.25)	0.76 (0.23 to 1.28)	0.70 (0.18 to 1.23)
TAU	0.54 (0.04 to 1.04)	0.73 (0.22 to 1.23)	N/A	0.76 (0.23 to 1.28)	0.70 (0.18 to 1.23)
<b>PBQ<sup>h</sup></b>					
ICBT	0.58 (0.05 to 1.11)	0.81 (0.27 to 1.36)	0.93 (0.38 to 1.49)	0.25 (-0.26 to 0.76)	0.23 (-0.28 to 0.74)
TAU	0.38 (-0.11 to 0.88)	0.57 (0.07 to 1.07)	N/A	0.25 (-0.26 to 0.76)	0.23 (-0.28 to 0.74)
<b>DAS-7<sup>i</sup></b>					
ICBT	0.37 (-0.16 to 0.90)	0.21 (-0.32 to 0.73)	0.47 (-0.06 to 1.00)	-0.36 (-0.87 to 0.15)	-0.58 (-1.10 to -0.07)
TAU	-0.09 (-0.58 to 0.40)	0.04 (-0.45 to 0.53)	N/A	-0.36 (-0.87 to 0.15)	-0.58 (-1.10 to -0.07)

<sup>a</sup>EPDS: Edinburgh Postnatal Depression Scale.

<sup>b</sup>ICBT: internet-delivered cognitive behavior therapy.

<sup>c</sup>TAU: treatment as usual.

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

<sup>e</sup>GAD-7: 7-item Generalized Anxiety Disorder.

<sup>f</sup>PHQ-9: 9-item Patient Health Questionnaire.

<sup>g</sup>DASS-21: 21-item Depression and Anxiety Stress Scale.

<sup>h</sup>PBQ: Postnatal Bonding Questionnaire.

<sup>i</sup>DAS-7: 7-item Dyadic Adjustment Scale.

## Secondary Outcomes

Means, SDs, and proportional reductions for all the secondary measures are shown in Table 2. The ICBT treatment group showed large improvements on the total DASS-21 and on the DASS-21 Anxiety and Stress subscales, large to medium improvements on the DASS-21 Depression subscale and the PBQ, and small improvements on the DAS-7 after treatment and at the 1- and 6-month follow-ups (Table 2). The TAU group had medium improvements on the DASS-21 and DASS-21 Stress subscale, medium to small improvements on the DASS-21 Anxiety subscale and the PBQ, small improvements on the DASS-21 Depression subscale, and negligible changes on the DAS-7. Between-group Cohen *d* effect sizes were medium on the DASS-21, DASS-21 Anxiety subscale, and DASS-21 Stress subscale; medium to small on the DAS-7; and small on the DASS-21 Depression subscale and the PBQ (Table 3). The between-group Cohen *d* effect sizes favored the ICBT treatment group on all measures except the DAS-7, where the TAU group had better scores after treatment and at the 1-month follow-up. However, interpretation of the DAS-7 is complicated by the fact that the groups had significantly different averages before treatment.

Hypothesis tests on the time×group interactions showed that the differences in proportional improvements were statistically significant for the DASS-21 ( $P=.04$ ), DASS-21 Stress subscale

( $P=.03$ ), and the DAS-7 after treatment ( $P=.04$ ) but not at the 1-month follow-up. Differences in proportional reductions were not significant after treatment or at the 1-month follow-up on the DASS-21 Anxiety subscale, DASS-21 Depression subscale, or the PBQ.

## Service Use

Service use after treatment and at the 1-month and 6-month follow-ups is summarized in Table 1. Between-group comparisons found no significant differences in health service use ( $P=.18-.99$ ). Comparisons were not possible at the 6-month follow-up as only treatment clients answered these questions.

## Treatment Adherence, Acceptability, and Satisfaction

Table 4 includes details about treatment adherence, acceptability, and satisfaction. In the ICBT treatment group, 75% (21/28) of the participants completed at least four of the five lessons, and 50% (14/28) of the participants completed all 5 lessons. Before treatment, the mean score on the credibility factor of the CEQ was 21.22 (SD 3.38), and the mean score on the expectancy factor was 17.07 (SD 3.77), with no significant differences between the ICBT treatment and TAU groups ( $P=.59$  and  $P=.44$  on the credibility and expectancy factors, respectively). Clients in the ICBT treatment group demonstrated a significant increase in treatment credibility scores after treatment (mean 23.58, SD 3.02;  $t_1=2.66$ ;  $P<.05$ ).

**Table 4.** Intervention use and treatment satisfaction (N=22).

Variables	Values
<b>Intervention use, mean (SD; range)</b>	
Number of log-ins	22.25 (13.81; 5-73)
Messages sent to the therapist	5.32 (5.23; 0-29)
Messages received from the therapist	9.89 (1.03; 8-12)
Number of phone calls from the therapist	1.93 (1.96; 0-6)
Furthest lesson accessed	4.57 (0.78; 2-5)
<b>Working alliance (WAI-SR<sup>a</sup>; mean [SD; range])</b>	
Total	47.77 (8.65; 22.00-58.00)
Goal subscale	15.36 (3.74; 5.00-20.00)
Task subscale	16.27 (2.00; 11.00-20.00)
Bond subscale	16.14 (3.67; 6.00-20.00)
<b>Treatment satisfaction, n (%)</b>	
Would recommend the course to a friend	22 (100)
The course was worth their time	22 (100)
Satisfied or very satisfied with the treatment	21 (95)
Satisfied or very satisfied with course materials	20 (91)
Increased or greatly increased confidence in managing symptoms	21 (95)
Increased or greatly increased motivation to seek additional treatment if needed in the future	17 (77)

<sup>a</sup>WAI-SR: Working Alliance Inventory-Short Revised.

## Discussion

### Principal Findings

In this study, 63 new mothers were randomized to receive either the *Wellbeing Course for New Moms* (ICBT) or TAU (ie, continued access to their standard care providers). Medium between-group Cohen *d* effects were observed in favor of the ICBT condition on most measures of anxiety and depression after treatment and at the follow-ups, with the exception of the depression subscale of the DASS-21. However, when examining the time-by-group interaction comparing differences in proportional reductions from before treatment between ICBT and TAU, the differences were only significant for the GAD-7 both after treatment and at the 1-month follow-up. Owing to the small sample size, this study was underpowered to detect significant differences in recovery rates between the treatment and TAU conditions, which may help explain the difference in findings based on effect sizes compared with time-by-group interactions of the proportional reductions. Improvements on all primary (EPDS, GAD-7, and PHQ-9) and most secondary measures (DASS-21 and PBQ) were maintained at the 6-month follow-up in the ICBT group. Finally, clients were satisfied with the ICBT course, perceived it as credible, and expressed high levels of alliance with their therapists.

The symptom improvement demonstrated in this sample is consistent with the treatment responses found in the face-to-face CBT literature. Meta-analyses have found statistically significant effects of CBT in the treatment of PPD, and improvements were maintained at the 6-month follow-up [37,38]. Although no known individual CBT studies have specifically investigated PPA, studies that report on comorbid anxiety symptoms [39,40] report that CBT is also effective in improving PPA symptoms. In terms of other ICBT trials, the *MUMentum Postnatal* program [18] is the only other published trial of transdiagnostic ICBT for the symptoms of PPD and PPA. In that trial, large between-group differences were found in favor of ICBT over TAU on measures of anxiety, depression, and distress after treatment and at follow-up. In this trial, medium between-group effects were found in favor of ICBT on all primary outcome measures (EPDS, PHQ-9, and GAD-7). Despite medium between-group effects, the differences in proportional improvements were only significant for the GAD-7 both after treatment and at the 1-month follow-up. Two notable differences between the trials included the proportion of clients who were referred by a medical or mental health professional (60% in this trial vs 6% in the study by Loughnan et al [18]) and the proportion of patients taking medication before treatment (40% in this trial vs 8% in the study by Loughnan et al [18]), although differences may have emerged because of the exclusion criteria for each study. Most clients in the study by Loughnan et al [18] self-referred to the program, suggesting that they may have been more motivated to complete the treatment, potentially resulting in better outcomes. Furthermore, the *MUMentum Postnatal* program contained more psychoeducation relevant to new mothers throughout the program (as opposed to a stand-alone resource) and is a briefer, more condensed treatment than the *Wellbeing Course for New Moms*, which may be more

appropriate for this population, given the time constraints of caring for a new infant.

### Strengths

Given the scarcity of published trials examining ICBT for PPA and PPD symptoms, this pilot study makes an important contribution to the literature [41]. A strength of this study is that it examined a transdiagnostic intervention as opposed to a disorder-specific intervention, which allows new mothers to address comorbid symptoms simultaneously instead of requiring separate ICBT courses. Furthermore, this trial provided helpful information about the trajectory of PPA and PPD symptoms among new mothers who were assigned to the TAU group. Specifically, new mothers in the TAU group reported significant improvements across primary measures, suggesting that symptoms of PPA and PPD may decrease in the absence of ICBT or perhaps that knowledge of upcoming treatment can facilitate symptom change.

### Limitations and Future Directions

Despite the contributions of this pilot study to the literature on ICBT for symptoms of PPA and PPD, this study had several limitations that should be considered when interpreting the results. Owing to the timing of the trial, approximately half of the clients were enrolled in the *Wellbeing Course for New Moms* during the COVID-19 pandemic; of note, research indicates that the COVID-19 pandemic has led to a substantial increase in the rates of PPD and PPA [42]. Clients commented on some of the challenges of engaging with ICBT during the pandemic and noted that opportunities to practice strategies (eg, pleasant activity scheduling or graded exposure) may have been limited by public health recommendations for physical distancing and self-isolation. It will be important to assess whether the findings from this study are replicated after the pandemic. An additional limitation of this pilot study was its small sample size, which meant that it was not reliably powered to detect small to medium between-group effects. Furthermore, the small sample size limited the types of analyses that could be conducted and prevented us from being able to determine moderators of symptom change (eg, relationship satisfaction, symptom severity and onset, use of medication, and access to other mental health supports). Future trials of the *Wellbeing Course for New Moms* should include larger sample sizes to ensure sufficient power for additional analyses. In future trials, it would be interesting to explore the extent to which individuals with PPA or PPD use additional resources. Finally, this study relied on self-report symptom measures as opposed to engaging participants in structured clinical interviews to ascertain information about their symptoms. Although this is not uncommon in the ICBT literature, it resulted in a lack of clarity regarding the participants' diagnostic status.

### Conclusions

This pilot study is the first known study of therapist-assisted, transdiagnostic ICBT designed to target symptoms of PPD and PPA. We found that the treatment resulted in larger improvements in depression, anxiety, and overall distress relative to a TAU condition when examining between-group effect sizes. However, the time-by-group interaction was

significant only for the GAD-7 both after treatment and at the 1-month follow-up. Improvements in primary outcomes in the ICBT group were maintained at the 6-month follow-up. This study provides initial evidence for transdiagnostic ICBT as a potentially accessible, effective, and acceptable treatment for symptoms of PPD and PPA.

## Acknowledgments

The authors would like to acknowledge the clients, community advisors, screeners, therapists, research staff, research associates, students, and web developers associated with the Online Therapy Unit at the University of Regina. The authors would specifically like to acknowledge Kelly Adlam, who served as the therapist for this trial. The Online Therapy Unit receives funding from the Saskatchewan Ministry of Health to provide internet-delivered cognitive behavior therapy services. The coauthors NT and BD are funded by the Australian Government to operate the national MindSpot Clinic. The funders had no involvement in the study design, collection, analysis, or interpretation of the data.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

CONSORT eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 817 KB - [formative\\_v6i9e37216\\_app1.pdf](#) ]

## References

1. Rai S, Pathak A, Sharma I. Postpartum psychiatric disorders: early diagnosis and management. *Indian J Psychiatry* 2015 Jul;57(Suppl 2):S216-S221 [FREE Full text] [doi: [10.4103/0019-5545.161481](#)] [Medline: [26330638](#)]
2. Maternal Mental Health in Canada, 2018/2019. Statistics Canada. 2019 Jun. URL: <https://www150.statcan.gc.ca/n1/daily-quotidien/190624/dq190624b-eng.htm> [accessed 2022-08-25]
3. Agius A, Xuereb RB, Carrick-Sen D, Sultana R, Rankin J. The co-existence of depression, anxiety and post-traumatic stress symptoms in the perinatal period: a systematic review. *Midwifery* 2016 May;36:70-79. [doi: [10.1016/j.midw.2016.02.013](#)] [Medline: [27106946](#)]
4. Field T. Postpartum depression effects on early interactions, parenting, and safety practices: a review. *Infant Behav Dev* 2010 Feb;33(1):1-6 [FREE Full text] [doi: [10.1016/j.infbeh.2009.10.005](#)] [Medline: [19962196](#)]
5. Field T. Postnatal anxiety prevalence, predictors and effects on development: a narrative review. *Infant Behav Dev* 2018 May;51:24-32. [doi: [10.1016/j.infbeh.2018.02.005](#)] [Medline: [29544195](#)]
6. Luca DL, Garlow N, Staatz C, Margiotta C, Zivin K. Societal costs of untreated perinatal mood and anxiety disorders in the United States. *Mathematica Policy Research*. 2019 Apr 29. URL: <https://www.mathematica.org/publications/societal-costs-of-untreated-perinatal-mood-and-anxiety-disorders-in-the-united-states> [accessed 2020-12-15]
7. Nillni YI, Mehralizade A, Mayer L, Milanovic S. Treatment of depression, anxiety, and trauma-related disorders during the perinatal period: a systematic review. *Clin Psychol Rev* 2018 Dec;66:136-148 [FREE Full text] [doi: [10.1016/j.cpr.2018.06.004](#)] [Medline: [29935979](#)]
8. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. *J Med Internet Res* 2016 Mar 07;18(3):e54 [FREE Full text] [doi: [10.2196/jmir.4993](#)] [Medline: [26952645](#)]
9. Goodman JH. Women's attitudes, preferences, and perceived barriers to treatment for perinatal depression. *Birth* 2009 Mar;36(1):60-69. [doi: [10.1111/j.1523-536X.2008.00296.x](#)] [Medline: [19278385](#)]
10. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. *World Psychiatry* 2019 Feb;18(1):20-28 [FREE Full text] [doi: [10.1002/wps.20610](#)] [Medline: [30600624](#)]
11. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cogn Behav Ther* 2018 Jan;47(1):1-18. [doi: [10.1080/16506073.2017.1401115](#)] [Medline: [29215315](#)]
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. Loughnan SA, Sie A, Hobbs MJ, Joubert AE, Smith J, Haskelberg H, et al. A randomized controlled trial of 'MUMentum Pregnancy': Internet-delivered cognitive behavioral therapy program for antenatal anxiety and depression. *J Affect Disord* 2019 Jan 15;243:381-390. [doi: [10.1016/j.jad.2018.09.057](#)] [Medline: [30266030](#)]
14. Forsell E, Bendix M, Holländare F, Szymanska von Schultz B, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: a randomised controlled trial. *J Affect Disord* 2017 Oct 15;221:56-64 [FREE Full text] [doi: [10.1016/j.jad.2017.06.013](#)] [Medline: [28628768](#)]

15. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. *Psychol Med* 2014 Jun;44(8):1675-1689 [FREE Full text] [doi: [10.1017/S0033291713002092](https://doi.org/10.1017/S0033291713002092)] [Medline: [24148703](https://pubmed.ncbi.nlm.nih.gov/24148703/)]
16. Pugh NE, Hadjistavropoulos HD, Dirkse D. A randomised controlled trial of therapist-assisted, Internet-delivered cognitive behavior therapy for women with maternal depression. *PLoS One* 2016 Mar 1;11(3):e0149186 [FREE Full text] [doi: [10.1371/journal.pone.0149186](https://doi.org/10.1371/journal.pone.0149186)] [Medline: [26930488](https://pubmed.ncbi.nlm.nih.gov/26930488/)]
17. Ashford MT, Olander EK, Rowe H, Fisher JR, Ayers S. Feasibility and acceptability of a Web-based treatment with telephone support for postpartum women with anxiety: randomized controlled trial. *JMIR Ment Health* 2018 Apr 20;5(2):e19 [FREE Full text] [doi: [10.2196/mental.9106](https://doi.org/10.2196/mental.9106)] [Medline: [29678804](https://pubmed.ncbi.nlm.nih.gov/29678804/)]
18. Loughnan SA, Butler C, Sie AA, Grierson AB, Chen AZ, Hobbs MJ, et al. A randomised controlled trial of 'MUMentum postnatal': Internet-delivered cognitive behavioural therapy for anxiety and depression in postpartum women. *Behav Res Ther* 2019 May;116:94-103. [doi: [10.1016/j.brat.2019.03.001](https://doi.org/10.1016/j.brat.2019.03.001)] [Medline: [30877878](https://pubmed.ncbi.nlm.nih.gov/30877878/)]
19. Hadjistavropoulos HD, Pugh NE, Nugent MM, Hesser H, Andersson G, Ivanov M, et al. Therapist-assisted Internet-delivered cognitive behavior therapy for depression and anxiety: translating evidence into clinical practice. *J Anxiety Disord* 2014 Dec;28(8):884-893 [FREE Full text] [doi: [10.1016/j.janxdis.2014.09.018](https://doi.org/10.1016/j.janxdis.2014.09.018)] [Medline: [25445078](https://pubmed.ncbi.nlm.nih.gov/25445078/)]
20. Hadjistavropoulos HD, Nugent MM, Alberts NM, Staples L, Dear BF, Titov N. Transdiagnostic Internet-delivered cognitive behaviour therapy in Canada: an open trial comparing results of a specialized online clinic and nonspecialized community clinics. *J Anxiety Disord* 2016 Aug;42:19-29 [FREE Full text] [doi: [10.1016/j.janxdis.2016.05.006](https://doi.org/10.1016/j.janxdis.2016.05.006)] [Medline: [27244278](https://pubmed.ncbi.nlm.nih.gov/27244278/)]
21. APA Presidential Task Force on Evidence-Based Practice. Evidence-based practice in psychology. *Am Psychol* 2006;61(4):271-285. [doi: [10.1037/0003-066X.61.4.271](https://doi.org/10.1037/0003-066X.61.4.271)] [Medline: [16719673](https://pubmed.ncbi.nlm.nih.gov/16719673/)]
22. 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](https://doi.org/10.1192/bjp.150.6.782)] [Medline: [3651732](https://pubmed.ncbi.nlm.nih.gov/3651732/)]
23. 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](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]
24. Titov N, Dear BF, Staples LG, Terides MD, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: a randomized controlled trial. *J Anxiety Disord* 2015 Oct;35:88-102 [FREE Full text] [doi: [10.1016/j.janxdis.2015.08.002](https://doi.org/10.1016/j.janxdis.2015.08.002)] [Medline: [26422822](https://pubmed.ncbi.nlm.nih.gov/26422822/)]
25. Hadjistavropoulos HD, Schneider LH, Klassen K, Dear BF, Titov N. Development and evaluation of a scale assessing therapist fidelity to guidelines for delivering therapist-assisted Internet-delivered cognitive behaviour therapy. *Cogn Behav Ther* 2018 Nov;47(6):447-461. [doi: [10.1080/16506073.2018.1457079](https://doi.org/10.1080/16506073.2018.1457079)] [Medline: [29693533](https://pubmed.ncbi.nlm.nih.gov/29693533/)]
26. Simpson W, Glazer M, Michalski N, Steiner M, Frey BN. Comparative efficacy of the generalized anxiety disorder 7-item scale and the Edinburgh Postnatal Depression Scale as screening tools for generalized anxiety disorder in pregnancy and the postpartum period. *Can J Psychiatry* 2014 Aug;59(8):434-440 [FREE Full text] [doi: [10.1177/070674371405900806](https://doi.org/10.1177/070674371405900806)] [Medline: [25161068](https://pubmed.ncbi.nlm.nih.gov/25161068/)]
27. 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](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
28. Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. *Ann Fam Med* 2009;7(1):63-70 [FREE Full text] [doi: [10.1370/afm.933](https://doi.org/10.1370/afm.933)] [Medline: [19139451](https://pubmed.ncbi.nlm.nih.gov/19139451/)]
29. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther* 1995 Mar;33(3):335-343. [doi: [10.1016/0005-7967\(94\)00075-u](https://doi.org/10.1016/0005-7967(94)00075-u)] [Medline: [7726811](https://pubmed.ncbi.nlm.nih.gov/7726811/)]
30. Brockington IF, Oates J, George S, Turner D, Vostanis P, Sullivan M, et al. A screening questionnaire for mother-infant bonding disorders. *Arch Womens Ment Health* 2001 Mar 15;3(4):133-140 [FREE Full text] [doi: [10.1007/s007370170010](https://doi.org/10.1007/s007370170010)]
31. Sharpley CF, Rogers HJ. Preliminary validation of the abbreviated Spanier Dyadic Adjustment Scale: some psychometric data regarding a screening test of marital adjustment. *Educ Psychol Meas* 1984 Dec 1;44(4):1045-1049. [doi: [10.1177/0013164484444029](https://doi.org/10.1177/0013164484444029)]
32. Dear BF, Titov N, Schwencke G, Andrews G, Johnston L, Craske MG, et al. An open trial of a brief transdiagnostic Internet treatment for anxiety and depression. *Behav Res Ther* 2011 Dec;49(12):830-837. [doi: [10.1016/j.brat.2011.09.007](https://doi.org/10.1016/j.brat.2011.09.007)] [Medline: [21993153](https://pubmed.ncbi.nlm.nih.gov/21993153/)]
33. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry* 2000 Jun;31(2):73-86. [doi: [10.1016/s0005-7916\(00\)00012-4](https://doi.org/10.1016/s0005-7916(00)00012-4)] [Medline: [11132119](https://pubmed.ncbi.nlm.nih.gov/11132119/)]
34. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clin Psychol Psychother* 2010;17(3):231-239. [doi: [10.1002/cpp.658](https://doi.org/10.1002/cpp.658)] [Medline: [20013760](https://pubmed.ncbi.nlm.nih.gov/20013760/)]
35. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986 Apr;73(1):13-22. [doi: [10.1093/biomet/73.1.13](https://doi.org/10.1093/biomet/73.1.13)]

36. Karin E, Crane MF, Dear BF, Nielssen O, Heller GZ, Kayrouz R, et al. Predictors, outcomes, and statistical solutions of missing cases in Web-based psychotherapy: methodological replication and elaboration study. *JMIR Ment Health* 2021 Feb 05;8(2):e22700 [FREE Full text] [doi: [10.2196/22700](https://doi.org/10.2196/22700)] [Medline: [33544080](https://pubmed.ncbi.nlm.nih.gov/33544080/)]
37. Haseli A, Mohammadi S. The effect of cognitive behavioral therapy on postpartum depression: a review and meta-analysis study. *Intern J Health Stud* 2019 Dec 28;5(4):15-20. [doi: [10.22100/ijhs.v5i4.709](https://doi.org/10.22100/ijhs.v5i4.709)]
38. Sockol LE. A systematic review of the efficacy of cognitive behavioral therapy for treating and preventing perinatal depression. *J Affect Disord* 2015 May 15;177:7-21. [doi: [10.1016/j.jad.2015.01.052](https://doi.org/10.1016/j.jad.2015.01.052)] [Medline: [25743368](https://pubmed.ncbi.nlm.nih.gov/25743368/)]
39. Milgrom J, Negri LM, Gemmill AW, McNeil M, Martin PR. A randomized controlled trial of psychological interventions for postnatal depression. *Br J Clin Psychol* 2005 Nov;44(Pt 4):529-542. [doi: [10.1348/014466505X34200](https://doi.org/10.1348/014466505X34200)] [Medline: [16368032](https://pubmed.ncbi.nlm.nih.gov/16368032/)]
40. Milgrom J, Gemmill AW, Ericksen J, Burrows G, Buist A, Reece J. Treatment of postnatal depression with cognitive behavioural therapy, sertraline and combination therapy: a randomised controlled trial. *Aust N Z J Psychiatry* 2015 Mar;49(3):236-245. [doi: [10.1177/0004867414565474](https://doi.org/10.1177/0004867414565474)] [Medline: [25586754](https://pubmed.ncbi.nlm.nih.gov/25586754/)]
41. Pearson N, Naylor PJ, Ashe MC, Fernandez M, Yoong SL, Wolfenden L. Guidance for conducting feasibility and pilot studies for implementation trials. *Pilot Feasibility Stud* 2020 Oct 31;6(1):167 [FREE Full text] [doi: [10.1186/s40814-020-00634-w](https://doi.org/10.1186/s40814-020-00634-w)] [Medline: [33292770](https://pubmed.ncbi.nlm.nih.gov/33292770/)]
42. Davenport MH, Meyer S, Meah VL, Strynadka MC, Khurana R. Moms are not OK: COVID-19 and maternal mental health. *Front Glob Womens Health* 2020 Jun 19;1:1 [FREE Full text] [doi: [10.3389/fgwh.2020.00001](https://doi.org/10.3389/fgwh.2020.00001)] [Medline: [34816146](https://pubmed.ncbi.nlm.nih.gov/34816146/)]

## Abbreviations

**CBT:** cognitive behavioral therapy  
**CEQ:** Credibility and Expectancy Questionnaire  
**DAS-7:** 7-item Dyadic Adjustment Scale  
**DASS-21:** 21-item Depression and Anxiety Stress Scale  
**EPDS:** Edinburgh Postnatal Depression Scale  
**GAD-7:** 7-item Generalized Anxiety Disorder  
**GEE:** generalized estimating equation  
**ICBT:** internet-delivered cognitive behavioral therapy  
**PBQ:** Postnatal Bonding Questionnaire  
**PHQ-9:** 9-item Patient Health Questionnaire  
**PPA:** postpartum anxiety  
**PPD:** postpartum depression  
**REDCap:** Research Electronic Data Capture  
**TAU:** treatment as usual  
**TSQ:** Treatment Satisfaction Questionnaire

*Edited by A Mavragani; submitted 11.02.22; peer-reviewed by A Gemmill, B Li, L Hantsoo; comments to author 20.07.22; revised version received 29.07.22; accepted 10.08.22; published 06.09.22.*

*Please cite as:*

Suchan V, Peynenburg V, Thiessen D, Nugent M, Dear B, Titov N, Hadjistavropoulos H  
*Transdiagnostic Internet-Delivered Cognitive Behavioral Therapy for Symptoms of Postpartum Anxiety and Depression: Feasibility Randomized Controlled Trial*  
*JMIR Form Res* 2022;6(9):e37216  
URL: <https://formative.jmir.org/2022/9/e37216>  
doi: [10.2196/37216](https://doi.org/10.2196/37216)  
PMID: [36066958](https://pubmed.ncbi.nlm.nih.gov/36066958/)

©Victoria Suchan, Vanessa Peynenburg, David Thiessen, Marcie Nugent, Blake Dear, Nickolai Titov, Heather Hadjistavropoulos. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 06.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# A Method to Deliver Automated and Tailored Intervention Content: 24-month Clinical Trial

Hailey N Miller<sup>1,2</sup>, PhD, RN; Corrine I Voils<sup>3,4</sup>, PhD; Kate A Cronin<sup>4</sup>, MPH; Elizabeth Jeanes<sup>4</sup>, MS; Jeffrey Hawley<sup>5</sup>, BS; Laura S Porter<sup>1,6</sup>, PhD; Rachel R Adler<sup>7</sup>, ScD, RD; Whitney Sharp<sup>3</sup>, MPH, RD; Samantha Pabich<sup>4,8</sup>, MD, MPH; Kara L Gavin<sup>3,4</sup>, PhD, MPH; Megan A Lewis<sup>9</sup>, PhD; Heather M Johnson<sup>10,11</sup>, MD, MS; William S Yancy Jr<sup>12</sup>, MD; Kristen E Gray<sup>13</sup>, PhD; Ryan J Shaw<sup>1</sup>, PhD, RN

<sup>1</sup>School of Nursing, Duke University, North Carolina, NC, United States

<sup>2</sup>Duke Global Digital Health Science Center, Duke University, Durham, NC, United States

<sup>3</sup>Department of Surgery, School of Medicine & Public Health, University of Wisconsin-Madison, Madison, WI, United States

<sup>4</sup>William S Middleton Memorial Veterans Hospital, Madison, WI, United States

<sup>5</sup>Duke Office of Clinical Research, School of Medicine, Duke University, Durham, NC, United States

<sup>6</sup>Department of Psychiatry & Behavioral Sciences, School of Medicine, Duke University, Durham, NC, United States

<sup>7</sup>Center for Surgery and Public Health, Brigham and Women's Hospital, Boston, MA, United States

<sup>8</sup>Department of Medicine, School of Medicine & Public Health, University of Wisconsin-Madison, Madison, WI, United States

<sup>9</sup>RTI International, Research Triangle Park, NC, United States

<sup>10</sup>Division of Cardiology, Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL, United States

<sup>11</sup>Christine E Lynn Women's Health & Wellness Institute/Baptist Health South Florida, Boca Raton, FL, United States

<sup>12</sup>Department of Medicine, Duke University School of Medicine, Durham, NC, United States

<sup>13</sup>Health Services Research & Development, Department of Veterans Affairs Puget Sound Health Care System, Seattle, WA, United States

**Corresponding Author:**

Ryan J Shaw, PhD, RN

School of Nursing

Duke University

307 Trent Drive

North Carolina, NC, 27710

United States

Phone: 1 919 684 9434

Email: [ryan.shaw@duke.edu](mailto:ryan.shaw@duke.edu)

## Abstract

**Background:** The use of digital technologies and software allows for new opportunities to communicate and engage with research participants over time. When software is coupled with automation, we can engage with research participants in a reliable and affordable manner. Research Electronic Data Capture (REDCap), a browser-based software, has the capability to send automated text messages. This feature can be used to automate delivery of tailored intervention content to research participants in interventions, offering the potential to reduce costs and improve accessibility and scalability.

**Objective:** This study aimed to describe the development and use of 2 REDCap databases to deliver automated intervention content and communication to index participants and their partners (dyads) in a 2-arm, 24-month weight management trial, Partner2Lose.

**Methods:** Partner2Lose randomized individuals with overweight or obesity and cohabitating with a partner to a weight management intervention alone or with their partner. Two databases were developed to correspond to 2 study phases: one for weight loss initiation and one for weight loss maintenance and reminders. The weight loss initiation database was programmed to send participants (in both arms) and their partners (partner-assisted arm) tailored text messages during months 1-6 of the intervention to reinforce class content and support goal achievement. The weight maintenance and reminder database was programmed to send maintenance-related text messages to each participant (both arms) and their partners (partner-assisted arm) during months 7-18. It was also programmed to send text messages to all participants and partners over the course of the 24-month trial to remind them of group classes, dietary recall and physical activity tracking for assessments, and measurement visits. All text messages were delivered via Twilio and were unidirectional.

**Results:** Five cohorts, comprising 231 couples, were consented and randomized in the Partner2Lose trial. The databases will send 53,518 automated, tailored text messages during the trial, significantly reducing the need for staff to send and manage intervention content over 24 months. The cost of text messaging will be approximately US \$450. Thus far, there is a 0.004% known error rate in text message delivery.

**Conclusions:** Our trial automated the delivery of tailored intervention content and communication using REDCap. The approach described provides a framework that can be used in future behavioral health interventions to create an accessible, reliable, and affordable method for intervention delivery and engagement that requires minimal trial-specific resources and personnel time.

**Trial Registration:** ClinicalTrials.gov NCT03801174; <https://clinicaltrials.gov/ct2/show/NCT03801174?term=NCT03801174>

(*JMIR Form Res* 2022;6(9):e38262) doi:[10.2196/38262](https://doi.org/10.2196/38262)

## KEYWORDS

text message; weight management; automation; clinical trial; engagement; digital technology; electronic data capture; REDCap; automated text message; digital health intervention; health intervention; health database; digital health

## Introduction

The use of digital technologies, such as smartphones, allows for new opportunities to consistently communicate and engage with research participants. With 85% of the adult US population now owning a smartphone [1], researchers can regularly send and receive data from participants in their everyday environment. This is particularly beneficial for researchers conducting behavioral health trials, as mobile technologies facilitate reaching participants in their day-to-day environments. This allows for cues to action or requests for information to reach participants when a specific behavior change is needed to achieve the desired outcome [2]. Further, when software are coupled with automation, researchers can engage with participants over a long period of time in an efficient and affordable manner [3,4]. This creates the opportunity for interventions to be scaled and reach populations across geographies.

The use of automated software to augment intervention content delivery offers several benefits. First, it can drastically reduce the need for personnel time. Using study staff to manually deliver intervention content to participants requires consistent staff effort. However, if software are leveraged, automated intervention content delivery can often reduce the amount of personnel time needed after initial programming [5]. Further, software can allow for intervention content to be tailored, which has shown to be an effective strategy for behavioral interventions [6,7]. The timing of delivery and information requests can be tailored to characteristics of the individual or population, such as an individual's behavior change goals, daily habits, or stage of behavior change, and in doing so, making the intervention content more relevant to the recipient [2]. For example, the frequency of contact can be programmed to be greater during the weight loss initiation stage and lower during the weight loss maintenance stage [8]. Similarly, the messaging content can be programmed to be different dependent on individualized goals, such as eating more fruits and vegetables or walking daily. This engagement can be expanded to support persons as well. Moreover, because software can maintain a record of the message delivery schedule, fidelity increases with automation. Thus, developing affordable and widely accessible strategies to communicate and engage with participants over time can improve intervention delivery.

Research Electronic Data Capture (REDCap), a browser-based software developed by Vanderbilt University for clinical research data collection and management, is increasingly available to academic institutions across the United States and in other countries, with over 5900 institutional partners in 145 countries [9,10]. REDCap has the capability to send automated text messages. Previous literature has described REDCap's ability to send automated communication, specifically in the context of automated medication reminders [11], symptom monitoring [12], postprocedure communication [13,14], and data collection surveys via text message over time [15]. It also has been previously used to support intervention workflow and content delivery [16-18]. However, the current literature does not include an in-depth report of REDCap being used to automate tailored intervention content and communication to research participants in a behavioral health clinical trial with longitudinal outcomes. Further, we are not aware of literature describing this approach to impart common or differentiated information between and within dyads. Given REDCap's broad use amongst academic institutions, and its affordability to research teams at institutions with licenses, it is an ideal platform to use for automating intervention communication in this context.

In this methods paper, we present a case study and lessons learned in which we use REDCap to deliver automated text messages to participants and their partners (dyads) in a weight management trial, Partner2Lose [19]. Our goal was to create an affordable, accessible method to communicate with research participants over time.

## Methods

### Study Design and Overview

Details of the Partner2Lose protocol have been published [19]. In brief, Partner2Lose was a parallel, 2-arm randomized controlled trial that carried out a comparative evaluation of a partner-assisted intervention and a participant-only intervention for weight management. The primary outcome was participant weight at 24 months. To be eligible to enroll, participants must have (1) been cohabitating with a partner, (2) had a BMI of 27-29.9 kg/m<sup>2</sup> and one obesity-related comorbidity or a BMI of ≥30 kg/m<sup>2</sup>, and (3) had a desire to lose weight. Partners had

to have a BMI  $\geq 18.5$  kg/m<sup>2</sup> to participate. Additional inclusion and exclusion criteria can be found in the previously published protocol.

Enrollment for Partner2Lose started in January 2019. Five cohorts of 45-50 couples were sequentially recruited and randomized to the participant-only arm or partner-assisted arm. The intervention included 6 months of weight loss initiation and 12 months of weight maintenance, followed by a 6-month period of no intervention. In the partner-assisted arm, partners participated in the intervention alongside the index participant and received communication skills training. Data collection will be completed in March 2023.

Index participants in both study arms received the nutrition, physical activity, and weight management intervention, a standard reduced-calorie weight management approach established in previous trials [20,21]. The weight loss initiation phase consisted of group classes every 2 weeks co-led by a registered dietitian and exercise physiologist and focused on unique dietary and physical activity education topics. At each class, participants were asked to select a goal topic from a menu

of 3-4 of them (see examples in [Table 1](#)) related to the class education topics. Participants provided their goal selection to a research staff member after the group meeting. These goal selections were used to inform tailored text messages to participants.

Partners in the partner-assisted arm also received communication and support skills training [22,23]. Partners attended half of the group meetings, where they received the same nutrition and activity education as index participants. The participants and their partners were asked to select a partner support plan to support the participants' goal from a prespecified list of options (see [Table 1](#)). Partners provided their support plan selection to a research staff member after the group meeting. These support plans were used to inform tailored text messages to partners.

During month 7, participants transitioned to the weight loss maintenance phase, where participants started to receive telephone support from the registered dietitian. During each of the 8 calls that were delivered, participants reflected on satisfaction with outcomes of weight loss and formed relapse prevention, self-monitoring, and social support plans. Partners in the partner-assisted joined 5 telephone calls during this period.

**Table 1.** Class schedule and associated goal selections for the Partner2Lose weight loss initiation phase.

Class schedule	Associated goals for participants	Associated support plan for partners
Class 1: Introduction to a low-calorie diet	— <sup>a</sup>	—
Class 2: Interpreting food labels and setting SMART (specific, measurable, attainable, relevant, and timebound) goals	<ul style="list-style-type: none"> <li>• Serving sizes</li> <li>• Meals and snacks</li> <li>• Calorie meal plan</li> <li>• Fiber</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check-in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>
Class 3: Tracking diet and activity	<ul style="list-style-type: none"> <li>• Planning</li> <li>• Measuring</li> <li>• Tracking</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>
Class 4: Grocery shopping	<ul style="list-style-type: none"> <li>• Shopping list</li> <li>• Healthy snacks</li> <li>• Whole grains</li> <li>• Produce</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>
Class 5: Meal planning	<ul style="list-style-type: none"> <li>• Meals</li> <li>• Snacks</li> <li>• Grocery lists</li> <li>• Recipes</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>
Class 6: Healthy cooking and modifying recipes	<ul style="list-style-type: none"> <li>• Dairy</li> <li>• Meat</li> <li>• Preparation</li> <li>• Healthy recipes</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>
Class 7: Dining out	<ul style="list-style-type: none"> <li>• Restaurant menus</li> <li>• Modifications</li> <li>• Planning</li> <li>• Meals for home</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>

Class schedule	Associated goals for participants	Associated support plan for partners
Class 8: Dining out advanced practice	<ul style="list-style-type: none"> <li>• Substitutions</li> <li>• Portion size</li> <li>• Temptations</li> <li>• On the side</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>
Class 9: Physical activity	<ul style="list-style-type: none"> <li>• Planning</li> <li>• Daily activity</li> <li>• Try something new</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>
Class 10: Eating more fruits and vegetables	<ul style="list-style-type: none"> <li>• Meals</li> <li>• Snacks</li> <li>• New produce</li> <li>• Preparation</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check-in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>
Class 11: Mindful eating	<ul style="list-style-type: none"> <li>• Distractions</li> <li>• Triggers</li> <li>• Strategies</li> <li>• Types of mindless eating</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>
Class 12: Emotional eating	<ul style="list-style-type: none"> <li>• Emotions</li> <li>• Triggers</li> <li>• High-risk foods</li> <li>• Coping strategies</li> </ul>	<ul style="list-style-type: none"> <li>• Do it together</li> <li>• Provide gentle reminders</li> <li>• Praise your partner</li> <li>• Remember the long game</li> <li>• Check in with your partner</li> <li>• Be mindful of how your choices affect your partner's goals</li> <li>• Talk with your partner to develop a support plan at home</li> </ul>

<sup>a</sup>None determined.

### Software Used to Schedule and Deliver Text Messages

To maximize the opportunity for scalability, it is beneficial to automate components of the intervention where appropriate. To that end, we used REDCap to automate the delivery of text message communications to participants and their partners for the duration of the intervention. Specifically, 2 databases were developed, which ran in parallel: a weight loss initiation database and a weight maintenance and reminder database. The choice to develop 2 databases was to simplify the database building processes, as the databases were used for different

purposes and thus had different programming requirements (described below).

The weight loss initiation database was used to reinforce class contents and support goal achievement during the weight loss initiation phase (months 1-6). The weight maintenance and reminder database was used to deliver a battery of automated reminders throughout the 24-month study period. It was also used to deliver automated text messages that concentrated on behavioral maintenance principles (eg, relapse prevention) during the weight maintenance phase (months 7-18). Both REDCap databases stored text message contents and

participant-level information, including role (ie, participant or partner), record ID, and cell phone number.

To specify text message content, conditions, and schedule, we used a feature in REDCap, known as “Automated Survey Invitations” (Figure 1). This feature allows messages to be sent, via email or text message, at a designated time (eg, immediately, in 2 hours, or in 14 days) after a set of prespecified conditions

are met. Conditions are specified from data fields in the respective REDCap database, such as participant role. An example condition for the goal database would be “[participant role=partner].” Under this condition, the messages would only be sent to someone if their documented role in REDCap was “partner.” Conditions used for the databases are described in greater detail below.

**Figure 1.** Automated Survey Invitations in REDCap used to send text messages related to goals in Partner2Lose.

The default for the Automated Survey Invitation feature in REDCap is for the messages to include a hyperlink to a survey for the recipient to complete. For this study, however, text messages were unidirectional and were not intended to solicit a response; hence, we did not want to send a hyperlink to a survey in our text message. Therefore, we used an open HTML comment to eliminate the hyperlinked survey feature in the body of the Automated Survey Invitation to make the system deliver a plain text message.

To deliver the text messages to participants, we used a third-party communication application integrated within REDCap, known as Twilio (Twilio Inc). Twilio is a cloud communication platform that uses web services application programming interfaces (APIs) to make and receive phone calls and text messages. To do so, a Twilio account, 34-digit string identifier (SID), authentication token (Auth Token), and purchased phone number are required. One Twilio phone number can be used for the duration of the study and for all participants. The Account SID and Auth Token act as the Twilio account’s username and password and are used to inform Twilio from which account the API requests are derived. As such, we created a Twilio account to retrieve an Account SID and Auth Token, which we entered into our REDCap databases. We disabled the “Request Inspector” feature on our Twilio account to ensure that the server did not store participant information.

Disabling this feature is a requirement by our institution to remain HIPAA (Health Insurance Portability and Accountability Act)-compliant.

## Development of Text Message Conditions

### *Weight Loss Initiation Database: Automated and Tailored Biweekly Messaging From Baseline to 6 Months*

The weight loss initiation database was programmed to send participants (both arms) and their partners (partner-assisted arm) tailored text messages in 2-week increments during the first 6 months of the intervention to reinforce class content and support goal achievement. As described previously, participants and their partners were prompted to choose a goal and support plan, respectively, following each group class and to provide their selections to a study team member. Participants’ selections, along with which class they attended (class number), were entered into REDCap by a study team member. This documentation triggered a 2-week battery of automated messages on a standard delay of 2, 4, 6, 8, 10, and 12 days (see example of the database in Figure 2). The participant’s role and class number were used as conditions to tailor the 2-week battery of messages for each participant (eg, [participant role=partner] AND [class number=class 5]). The goal and support plan selections were piped into text messages, when relevant, to individualize text message content (see Table 2 for example).

Figure 2. Partner2Lose weight loss initiation database.

Instrument name	Fields	View PDF	Enabled as survey	Instrument actions	Survey-related options
person_info_and_messages	318			Choose action	Survey settings + Automated Invitations
P2l Po Sms Sender Class 01	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 02	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 03	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 04	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 05	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 06	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 07	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 08	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 09	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 10	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 11	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 12	1			Choose action	Survey settings Automated Invitations
P2l Po Sms Sender Class 13	1			Choose action	Survey settings Automated Invitations

**Automated Invitations**

Choose an event below for which to set up or modify automated survey invitations.

- Modify Message\_01
- Modify Message\_02
- Modify Message\_03
- Modify Message\_04
- Modify Message\_05
- Modify Message\_06

**Table 2.** Text message algorithm example for class 2: interpreting food labels and setting SMART (specific, measurable, attainable, relevant, and timebound) goals.

Participant role	Class schedule	Goal selection	Text message battery
Participant in participant-only arm	Class 2: Interpreting food labels and setting SMART goals	<ul style="list-style-type: none"> <li>Participant goal: meals and snacks</li> <li>Partner goal: not applicable</li> </ul>	<ul style="list-style-type: none"> <li>Day 2: Your goal for the next two weeks focuses on: Meals and snacks [populated from goal chosen in Table 1 from a menu of 3-4 options]</li> <li>Day 4: Remember to look at the nutrition facts label and stick to 1 serving to stay within your calorie budget</li> <li>Day 6: Keep an eye on added sugars</li> <li>Day 8: Remember to work on the goal you set for: Meals and snacks [populated from goal chosen in Table 1 from a menu of 3-4 options]</li> <li>Day 10: How much fiber have you been consuming? Fiber helps us stay full longer. Try for at least 25g/day</li> <li>Day 12: Try comparing the serving sizes of food using volume (cup measures) vs. weight (food scale). How do they compare?</li> </ul>
Participant in partner-assisted arm	Class 2: Interpreting food labels and setting SMART goals	<ul style="list-style-type: none"> <li>Participant goal: meals and snacks</li> <li>Partner goal: praise your partner</li> </ul>	<ul style="list-style-type: none"> <li>Day 2: Text 1: Your goal for the next two weeks focuses on: Meals and snacks. Text 2: Your partner's support plan is: Praise your partner</li> <li>Day 4: Support tip: When trying to problem-solve together: state the issue, say why it's important, discuss possible solutions, and try a solution.</li> <li>Day 6: Keep an eye on added sugars.</li> <li>Day 8: Text 1: Check in with your partner about whether their support plan is helping you reach the goal you set for: Meals and snacks. Text 2: Check in with your partner about how their support plan is going, which is: Praise your partner</li> <li>Day 10: Support tip: When making decisions with your partner, remember to talk about what each person needs out of the solution.</li> <li>Day 12: Try comparing the serving sizes of food using volume (cup measures) vs. weight (food scale). How do they compare?</li> </ul>
Partner in partner-assisted arm	Class 2: Interpreting food labels and setting SMART goals	<ul style="list-style-type: none"> <li>Participant goal: meals and snacks</li> <li>Partner goal: praise your partner</li> </ul>	<ul style="list-style-type: none"> <li>Day 2: Text 1: Your partner's goal for the next two weeks focuses on: Meals and Snacks. Text 2: Your support plan is: Praise your partner</li> <li>Day 4: Support tip: When trying to problem-solve together: state the issue, say why it's important, discuss possible solutions, and try a solution</li> <li>Day 6: Keep an eye on added sugars</li> <li>Day 8: Text 1: Check in with your partner about their goal, which focuses on: Meals and Snacks. Text 2: Check in with your partner about how you are doing with your support plan, which is: Praise your partner</li> <li>Day 10: Support tip: When making decisions with your partner, remember to talk about what each person needs out of the solution</li> <li>Day 12: Try comparing the serving sizes of food using volume (cup measures) vs. weight (food scale). How do they compare?</li> </ul>

**Maintenance and Reminder Database: Automated Messaging From Baseline to 24 Months**

The weight maintenance and reminder database was programmed to send a total of 80 automated text messages to each participant (both arms) and 70 automated text messages to their partners (partner-assisted arm) over the 24-month study period. These text messages were used to remind participants and their partners of group classes, dietary recall and physical activity tracking for assessments, and measurement visits. They were also used to reinforce behavior maintenance principles during the weight loss maintenance phase of the study (months 7-18). A REDCap field was created for each possible text message at baseline. In each field, we specified the text message

conditions and delivery schedule and piped in the corresponding text message content. Then, we set conditions to specify which participants should receive the text message. Conditions used in the 2-year database were the role and number of days since baseline (eg, [participant role=partner] AND [days since baseline=80]). To customize delivery dates for each cohort, we used an external date calculation module (date calculation fields, version 1.64) in REDCap. This individualization took into consideration the schedule of group meetings.

**Cost of Text Messaging**

Twilio has a pay-as-you-go payment structure and does not require an upfront fee to make an account. A monthly fee of US \$1.00 for a Twilio phone number is required. Phone numbers



can be selected on the basis of a local area code. At the time of this intervention, text messaging to domestic numbers costs US \$0.0075 per text message.

### Intervention Fidelity

REDCap automatically maintains a log display or comprehensive list of messages that are delivered or scheduled to be delivered (Figure 3). The log displays the date and time when the message was sent or is scheduled to be sent, as well

as the participant's cell phone number and record ID. The log also displays if there were any errors. The log can be viewed and downloaded at any time to monitor intervention fidelity. Intervention fidelity is further enhanced over long periods of time through automation and limiting the need for a person to send a message. Our study team looked at the log monthly to check for errors and confirm message delivery. If an error was seen, we investigated and resolved it.

**Figure 3.** Partner2Lose text message delivery log in Research Electronic Data Capture (REDCap).

Invitation send time	View Invite	Participant Email	Participant Phone	Record	Participant Identifier	Survey	Survey Link	Responded?	Errors (if any)	
04/12/2021 9:28am		[No email listed]		<a href="#">02574</a>		P2I PAR Sms Sender day_210_2				<input type="checkbox"/>
04/12/2021 9:28am		[No email listed]		<a href="#">02624</a>		P2I PAR Sms Sender day_210_2				<input type="checkbox"/>
04/12/2021 9:28am		[No email listed]		<a href="#">02644</a>		P2I PAR Sms Sender day_210_2				<input type="checkbox"/>
04/12/2021 9:28am		[No email listed]		<a href="#">02654</a>		P2I PAR Sms Sender day_210_2				<input type="checkbox"/>
04/12/2021 9:28am		[No email listed]		<a href="#">02664</a>		P2I PAR Sms Sender day_210_2				<input type="checkbox"/>
04/12/2021 9:28am		[No email listed]		<a href="#">02724</a>		P2I PAR Sms Sender day_210_2				<input type="checkbox"/>
04/12/2021 9:28am		[No email listed]		<a href="#">02734</a>		P2I PAR Sms Sender day_210_2				<input type="checkbox"/>

### Ethics Approval

Approval for Partner2Lose was obtained by the University of Wisconsin Health Sciences Institutional Review Board in December 2018 (protocol 2018-1400).

### Results

A total of 1061 participants and partner dyads were screened to participate in Partner2Lose. Among these dyads, 3 participants were excluded from the study for not having a smartphone with a data or texting plan. No partners were excluded for this reason. Five cohorts comprising 231 couples were consented and randomized in the Partner2Lose trial. The initial build of the 2 databases took approximately 120 hours. Thereafter, the databases have required, on average, less than 15 minutes of maintenance each week.

As of February 22, 2022, the initiation and maintenance and reminder databases have sent 22,887 and 21,447 Automated Survey Invitations via text message, respectively. A total of 53,518 text messages are planned to be delivered by study completion (March 2023).

The cost to send these text messages via Twilio will be approximately US \$450 for the full study duration. Of the 44,334 text messages that have been sent, 217 failed to deliver, equating to a 0.004% error rate. These errors appear to occur at random and consistently over time. REDCap did not provide more information regarding the messages that failed to deliver, which could indicate error on the participants' side, which we were unable to identify. In one instance, though, study staff reported

that participants said they received multiple repeated messages. This was a bug found within the REDCap logic, and we were able to quickly address the problem. Additional errors have not been noted within REDCap or from study participants thus far.

### Discussion

#### Principal Findings

To effectively implement interventions in large populations, scalable intervention delivery approaches are needed. This study demonstrates a feasible approach to automate a significant component of participant communication and engagement over a 24-month behavioral weight management trial using a widely available software, REDCap.

The widespread use of mobile technology has expanded our opportunity to automate the delivery of intervention content without requiring face-to-face interactions and with minimal upkeep and maintenance by study staff [1]. Inasmuch, it reduces study burden and limits the need to travel to a study site. This is particularly useful for populations in rural areas or those who may work multiple jobs or have caregiving responsibilities. Text messaging is also cost effective, as indicated in our report and previous research [24], and unlike other digital interventions and strategies, it does not require a smartphone (with access to applications) or other devices, such as wearables, facilitating reach to sociodemographically diverse groups [1,25]. In fact, 97% of adults own a cellphone in the United States, with high rates across all race and ethnicity, household income, and educational attainment groups [1].

There are other examples of automated, tailored interventions being delivered in behavioral health trials. These interventions have used other software tools that require individual development and may involve smartphone apps. One example is the Nourish study [26] that leverages a smartphone application and evidence-based behavior change principles to improve adherence to the DASH eating pattern among adults with hypertension. The Nourish intervention delivers tailored and personalized feedback that is based on dietary tracking and goal attainment. This feedback also takes into consideration personal characteristics of the participant, such as working full time, to inform the feedback provided. Given the increased number of variables being considered in the feedback algorithm, in addition to it relying on an external application, REDCap likely would not be a feasible option for this study or others alike. These are important factors to consider when selecting a software platform. We chose REDCap for this study owing to institutional investment in which it supports a wide variety of research needs, and REDCap's wide adoption and collaborative, international network. We also only sought to leverage text messaging for intervention delivery owing to its simplicity and equity in use across diverse populations. Thus, results from this study can be more easily replicated.

The REDCap Consortium is an international community of software platform partners [10]. While each institution maintains its own version of the software, developed features are shared across the community. This allows for interoperability of data transfer across sites and best practices related to security and regulatory requirements. The initial costs of using software such as REDCap requires institutional investment; however, there is less of a need to rebuild custom software tools across the enterprise once the infrastructure exists, resulting in reduced long-term spending [27]. Using REDCap, or a similar platform, is also cost savings for individual study teams, as the scalability of automated processes is exponential and more affordable than manual processes that incur labor costs over time. For example, increasing a sample size from 100 to 10,000 would traditionally require a significant increase in labor costs and data tracking; however, software and automation can greatly reduce the per-participant cost.

Automating the participant communication component of our intervention delivery increased our ability to monitor and manage intervention fidelity. Specifically, using REDCap's survey invitation log feature, we were able to identify messages delivered or scheduled to be delivered that had an error. This feature allowed us to make informed and responsive modifications such as reinitiating messages or stopping double messaging. It should be noted that our intervention did not ask or prompt participants to respond to text messages, so there could be delivery errors not identified by REDCap or our team. Nonetheless, the databases used in this study demonstrated a high rate of fidelity over time in message delivery as demonstrated by the very low error rate.

The tools and features in REDCap are regularly modified and updated to improve the platform's functionality and effectiveness. As such, the workflow described in this paper would benefit from being adapted to leverage the tools and features available today. For example, if we were to implement

automated text messaging in our next trial, we would use REDCap's "Alerts and Notifications" feature to send the text messages via Twilio. This relatively newer feature provides an option to specify delivery conditions to tailor the text message content by each participant, however, does not require the text message to be connected to a survey in REDCap. The feature described to send text messages in this report, Automated Survey Invitations, would be more appropriate to send text messages or communications that solicit a survey response, such as an outcome survey or adverse event questionnaire. Nonetheless, both of these features have tracking logs, in which intervention delivery and survey completion can be tracked and maintained by the study team.

### Limitations

We note several limitations to automating intervention content and communication using REDCap and Twilio, which should be considered before design and implementation. First, it requires participants to have a consistent mobile phone, which may limit the generalizability of the automated communication approach. Although mobile phone ownership is high and continues to grow in the United States [1], there is still a small percentage of individuals who do not own a mobile phone and thus cannot be reached using this approach. For example, 8% of adults aged 65 years or older do not own mobile phone and would be inaccessible via this modality [1]. Second, the phone number that is used to deliver the text messages via Twilio is a robot. Thus, if someone were to reply to the text message, there would not be a study team member monitoring the responses or be available to reply. This can be a safety concern if a participant is seeking medical advice from a study team member; however, to minimize this concern, we notified participants at the start of the trial that responses would not be monitored. Additionally, it is worth noting that participants received messages from the same phone number over the course of the study and were able to save the number so they would not think it was spam. Twilio has the capability for 2-way SMS text messaging; however, at the time of implementation, REDCap only had the capability of receiving text messages that met predefined criteria. For safety or technical concerns, we provided alternative contact modalities for the participants to reach the trial staff, including an email ID and a phone number. Third, although our data indicate that the error rate was small in this study, it is possible that errors went unidentified as our intervention did not prompt a response from participants. Moreover, if there is a programming error, it is likely to impact several study participants rather than a single participant. Fourth, given the use of text messaging, it was not possible to record if a participant read the intervention communication. Lastly, accounting for holidays and weather can be challenging, and automated messaging may occur when not desired by participants. In the future, this may be mitigated by asking participants communication preferences or tailoring text messages to times of the day that are best suited to them.

### Conclusions

The proliferation of mobile phones coupled with research management software offers study teams the opportunity to integrate automation at scale into their intervention

implementation and delivery. Our trial automated the delivery of intervention content and communications using REDCap for 5 cohorts of participants and their partners in a behavioral health trial with longitudinal outcomes. Our approach provides a

framework that can be used in future behavioral health interventions to create an affordable, reliable and accessible method for intervention engagement.

## Acknowledgments

This study was supported by a grant to CIV from the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK111491-01). CIV is also supported by a Research Career Scientist award (RCS 14-443) from the Health Services Research & Development Service of the Department of Veterans Affairs. Research reported in this publication was also supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (award UL1TR002553). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US government.

## Conflicts of Interest

None declared.

## References

1. Mobile Fact Sheet. Pew Research Center. 2021 Apr 07. URL: <https://www.pewresearch.org/internet/fact-sheet/mobile/> [accessed 2021-06-21]
2. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAI) 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](https://doi.org/10.1007/s12160-016-9830-8)] [Medline: [27663578](https://pubmed.ncbi.nlm.nih.gov/27663578/)]
3. Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). *Addiction* 2017 Jul 02;112(7):1238-1249 [FREE Full text] [doi: [10.1111/add.13802](https://doi.org/10.1111/add.13802)] [Medline: [28239919](https://pubmed.ncbi.nlm.nih.gov/28239919/)]
4. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. *Am J Prev Med* 2017 Mar;52(3):391-402. [doi: [10.1016/j.amepre.2016.10.042](https://doi.org/10.1016/j.amepre.2016.10.042)] [Medline: [28073656](https://pubmed.ncbi.nlm.nih.gov/28073656/)]
5. Marks RG, Conlon M, Ruberg SJ. Paradigm shifts in clinical trials enabled by information technology. *Stat Med* 2001 Sep 15;20(17-18):2683-2696. [doi: [10.1002/sim.736](https://doi.org/10.1002/sim.736)] [Medline: [11523076](https://pubmed.ncbi.nlm.nih.gov/11523076/)]
6. 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](https://doi.org/10.1080/10810730.2013.768727)] [Medline: [23750972](https://pubmed.ncbi.nlm.nih.gov/23750972/)]
7. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. *Digit Health* 2019 Feb 05;5:2055207619826685 [FREE Full text] [doi: [10.1177/2055207619826685](https://doi.org/10.1177/2055207619826685)] [Medline: [30783535](https://pubmed.ncbi.nlm.nih.gov/30783535/)]
8. Shaw RJ, Bosworth HB, Silva SS, Lipkus IM, Davis LL, Sha RS, et al. Mobile health messages help sustain recent weight loss. *Am J Med* 2013 Nov;126(11):1002-1009 [FREE Full text] [doi: [10.1016/j.amjmed.2013.07.001](https://doi.org/10.1016/j.amjmed.2013.07.001)] [Medline: [24050486](https://pubmed.ncbi.nlm.nih.gov/24050486/)]
9. 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](https://doi.org/10.1016/j.jbi.2008.08.010)] [Medline: [18929686](https://pubmed.ncbi.nlm.nih.gov/18929686/)]
10. 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](https://doi.org/10.1016/j.jbi.2019.103208)] [Medline: [31078660](https://pubmed.ncbi.nlm.nih.gov/31078660/)]
11. Pernell B, DeBaun M, Becker K, Rodeghier M, Bryant V, Cronin R. Improving medication adherence with two-way short message service reminders in sickle cell disease and asthma. *Appl Clin Inform* 2017 Dec 21;08(02):541-559. [doi: [10.4338/aci-2016-12-ra-0203](https://doi.org/10.4338/aci-2016-12-ra-0203)] [Medline: [28536723](https://pubmed.ncbi.nlm.nih.gov/28536723/)]
12. Naughton MJ, Salani R, Peng J, Lustberg M, DeGraffinreid C, Moon J, et al. Feasibility of implementing a text-based symptom-monitoring program of endometrial, ovarian, and breast cancer patients during treatment. *Qual Life Res* 2021 Nov 14;30(11):3241-3254 [FREE Full text] [doi: [10.1007/s11136-020-02660-w](https://doi.org/10.1007/s11136-020-02660-w)] [Medline: [33052514](https://pubmed.ncbi.nlm.nih.gov/33052514/)]
13. Gee KM, Jones RE, Nevarez N, McClain LE, Wools G, Beres AL. No pain is gain: A prospective evaluation of strict non-opioid pain control after pediatric appendectomy. *J Pediatr Surg* 2020 Jun;55(6):1043-1047. [doi: [10.1016/j.jpedsurg.2020.02.051](https://doi.org/10.1016/j.jpedsurg.2020.02.051)] [Medline: [32171535](https://pubmed.ncbi.nlm.nih.gov/32171535/)]
14. Gessner D, Hunter OO, Kou A, Mariano ER. Automated text messaging follow-up for patients who receive peripheral nerve blocks. *Reg Anesth Pain Med* 2021 Jun 01;46(6):524-528. [doi: [10.1136/rapm-2021-102472](https://doi.org/10.1136/rapm-2021-102472)] [Medline: [33649155](https://pubmed.ncbi.nlm.nih.gov/33649155/)]
15. Polillo A, Voineskos AN, Foussias G, Kidd SA, Sav A, Hawley S, et al. Using digital tools to engage patients with psychosis and their families in research: survey recruitment and completion in an early psychosis intervention program. *JMIR Ment Health* 2021 May 31;8(5):e24567 [FREE Full text] [doi: [10.2196/24567](https://doi.org/10.2196/24567)] [Medline: [34057421](https://pubmed.ncbi.nlm.nih.gov/34057421/)]

16. Hopkins CM, Miller HN, Brooks TL, Mo-Hunter L, Steinberg DM, Bennett GG. Designing ruby: protocol for a 2-Arm, brief, digital randomized controlled trial for internalized weight bias. *JMIR Res Protoc* 2021 Nov 25;10(11):e31307 [FREE Full text] [doi: [10.2196/31307](https://doi.org/10.2196/31307)] [Medline: [34842549](https://pubmed.ncbi.nlm.nih.gov/34842549/)]
17. Kianersi S, Luetke M, Ludema C, Valenzuela A, Rosenberg M. Use of research electronic data capture (REDCap) in a COVID-19 randomized controlled trial: a practical example. *BMC Med Res Methodol* 2021 Aug 21;21(1):175 [FREE Full text] [doi: [10.1186/s12874-021-01362-2](https://doi.org/10.1186/s12874-021-01362-2)] [Medline: [34418958](https://pubmed.ncbi.nlm.nih.gov/34418958/)]
18. Crane S, Comer RS, Arenson AD, Draucker C. Using REDCap to facilitate web-based therapeutic intervention research. *Nurs Res* 2019;68(6):483-487 [FREE Full text] [doi: [10.1097/NNR.0000000000000367](https://doi.org/10.1097/NNR.0000000000000367)] [Medline: [31693554](https://pubmed.ncbi.nlm.nih.gov/31693554/)]
19. Voils CI, Shaw R, Adler R, Jeanes E, Lewis MA, Sharp W, et al. Protocol for Partner2Lose: A randomized controlled trial to evaluate partner involvement on long-term weight loss. *Contemp Clin Trials* 2020 Sep;96:106092 [FREE Full text] [doi: [10.1016/j.cct.2020.106092](https://doi.org/10.1016/j.cct.2020.106092)] [Medline: [32750431](https://pubmed.ncbi.nlm.nih.gov/32750431/)]
20. Voils CI, Olsen MK, Gierisch JM, McVay MA, Grubber JM, Gaillard L, et al. Maintenance of weight loss after initiation of nutrition training. *Ann Intern Med* 2017 Feb 21;166(7):463. [doi: [10.7326/m16-2160](https://doi.org/10.7326/m16-2160)]
21. Yancy WS, Westman EC, McDuffie JR, Grambow SC, Jeffreys AS, Bolton J, et al. A randomized trial of a low-carbohydrate diet vs orlistat plus a low-fat diet for weight loss. *Arch Intern Med* 2010 Jan 25;170(2):136-145. [doi: [10.1001/archinternmed.2009.492](https://doi.org/10.1001/archinternmed.2009.492)] [Medline: [20101008](https://pubmed.ncbi.nlm.nih.gov/20101008/)]
22. Epstein N, Baucom DH. *Enhanced Cognitive-behavioral Therapy for Couples: A Contextual Approach*. Washington, DC: American Psychological Association; 2002.
23. Lewis MA, McBride CM, Pollak KI, Puleo E, Butterfield RM, Emmons KM. Understanding health behavior change among couples: an interdependence and communal coping approach. *Soc Sci Med* 2006 Mar;62(6):1369-1380. [doi: [10.1016/j.socscimed.2005.08.006](https://doi.org/10.1016/j.socscimed.2005.08.006)] [Medline: [16146666](https://pubmed.ncbi.nlm.nih.gov/16146666/)]
24. Cobos-Campos R, Mar J, Apiñaniz A, de Lafuente AS, Parraza N, Aizpuru F, et al. Cost-effectiveness analysis of text messaging to support health advice for smoking cessation. *Cost Eff Resour Alloc* 2021 Feb 15;19(1):9 [FREE Full text] [doi: [10.1186/s12962-021-00262-y](https://doi.org/10.1186/s12962-021-00262-y)] [Medline: [33588885](https://pubmed.ncbi.nlm.nih.gov/33588885/)]
25. Willcox JC, Dobson R, Whittaker R. Old-fashioned technology in the era of "bling": is there a future for text messaging in health care? *J Med Internet Res* 2019 Dec 20;21(12):e16630 [FREE Full text] [doi: [10.2196/16630](https://doi.org/10.2196/16630)] [Medline: [31859678](https://pubmed.ncbi.nlm.nih.gov/31859678/)]
26. Miller HN, Berger MB, Askew S, Kay MC, Hopkins CM, Iragavarapu MS, et al. The Nourish Protocol: A digital health randomized controlled trial to promote the DASH eating pattern among adults with hypertension. *Contemp Clin Trials* 2021 Oct;109:106539 [FREE Full text] [doi: [10.1016/j.cct.2021.106539](https://doi.org/10.1016/j.cct.2021.106539)] [Medline: [34400362](https://pubmed.ncbi.nlm.nih.gov/34400362/)]
27. Klipin M, Mare I, Hazelhurst S, Kramer B. The process of installing REDCap, a web based database supporting biomedical research. *Appl Clin Inform* 2017 Dec 19;05(04):916-929. [doi: [10.4338/aci-2014-06-cr-0054](https://doi.org/10.4338/aci-2014-06-cr-0054)] [Medline: [25589907](https://pubmed.ncbi.nlm.nih.gov/25589907/)]

## Abbreviations

**API:** application programming interface

**Auth:** authentication

**HIPAA:** Health Insurance Portability and Accountability Act

**REDCap:** Research Electronic Data Capture

**SID:** string identifier

**SMART:** specific, measurable, attainable, relevant, and timebound

*Edited by T Leung, B Astaneh; submitted 25.03.22; peer-reviewed by M Patel, H Belli, M Kim, M Bardus; comments to author 23.06.22; revised version received 28.07.22; accepted 10.08.22; published 06.09.22.*

### *Please cite as:*

Miller HN, Voils CI, Cronin KA, Jeanes E, Hawley J, Porter LS, Adler RR, Sharp W, Pabich S, Gavin KL, Lewis MA, Johnson HM, Yancy Jr WS, Gray KE, Shaw RJ

*A Method to Deliver Automated and Tailored Intervention Content: 24-month Clinical Trial*

*JMIR Form Res* 2022;6(9):e38262

URL: <https://formative.jmir.org/2022/9/e38262>

doi: [10.2196/38262](https://doi.org/10.2196/38262)

PMID: [36066936](https://pubmed.ncbi.nlm.nih.gov/36066936/)

©Hailey N Miller, Corrine I Voils, Kate A Cronin, Elizabeth Jeanes, Jeffrey Hawley, Laura S Porter, Rachel R Adler, Whitney Sharp, Samantha Pabich, Kara L Gavin, Megan A Lewis, Heather M Johnson, William S Yancy Jr, Kristen E Gray, Ryan J Shaw. Originally published in *JMIR Formative Research* (<https://formative.jmir.org/>), 06.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits

unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Design, Development, and Testing of BEST4Baby, an mHealth Technology to Support Exclusive Breastfeeding in India: Pilot Study

Tony Ma<sup>1</sup>, MS; Katie Chang<sup>1</sup>, MS; Amal Alyusuf<sup>1</sup>, MD; Elina Bajracharya<sup>1</sup>, MS; Yukiko Washio<sup>2</sup>, PhD; Patricia J Kelly<sup>3</sup>, PhD; Roopa M Bellad<sup>4</sup>, MD; Niranjana S Mahantashetti<sup>4</sup>, MD; Umesh Charantimath<sup>4</sup>, MD; Vanessa L Short<sup>3</sup>, PhD; Parth Lalakia<sup>3</sup>, MPH; Frances Jaeger<sup>3</sup>, DrPH; Shivaprasad Goudar<sup>4</sup>, MD; Richard Derman<sup>3</sup>, MD

<sup>1</sup>Benten Technologies, Inc, Manassas, VA, United States

<sup>2</sup>Substance Use, Gender and Applied Research, RTI International, Research Triangle Park, NC, United States

<sup>3</sup>Thomas Jefferson University, Philadelphia, PA, United States

<sup>4</sup>KLE Academy of Higher Education and Research, Jawaharlal Nehru Medical College, Belagavi, Karnataka, India

**Corresponding Author:**

Tony Ma, MS

Benten Technologies, Inc

9408 Grant Avenue

Suite 206

Manassas, VA, 20110-1816

United States

Phone: 1 703 662 5858 ext 143

Fax: 1 703 560 1908

Email: [tonyma@bententech.com](mailto:tonyma@bententech.com)

## Abstract

**Background:** Exclusive breastfeeding (EBF) at 6 months of age in most low- and middle-income countries, including India, is surprisingly low. There is a relative lack of mobile health apps that specifically focus on leveraging the use of peer counselors (PCs) to support mothers as a means of increasing EBF practices in low- and middle-income countries.

**Objective:** This study aimed to design, develop, and test the usability of Breastfeeding Education Support Tool for Baby (BEST4Baby), a mobile health app specifically designed to support PCs in providing in-home breastfeeding counseling support to mothers in rural India on optimal breastfeeding practices.

**Methods:** A user-centered design process with an agile development methodology was used. The approach involved stakeholders and mothers who were trained to serve as PCs to guide BEST4Baby's design and development, including the app's content and features. PCs were engaged through focus groups with interactive wireframes. During the 24-month pilot study period, we conducted a feasibility test of the BEST4Baby app with 22 PCs who supported home visits with mothers residing in rural India. The intervention protocol required PCs to provide education and follow mothers using the BEST4Baby app, with 9 scheduled home visits from the late prenatal stage to 6 months post partum. BEST4Baby's usability from the PCs' perspective was assessed using the translated System Usability Scale (SUS).

**Results:** The findings of this study align with best practices in user-centered design (ie, understanding user experience, including context with iterative design with stakeholders) to address EBF barriers. This led to the cultural tailoring and contextual alignment of an evidence-based World Health Organization breastfeeding program with an iterative design and agile development of the BEST4Baby app. A total of 22 PCs tested and rated the BEST4Baby app as highly usable, with a mean SUS score of 85.3 (SD 9.1), placing it over the 95th percentile for SUS scores. The approach translated into a highly usable BEST4Baby app for use by PCs in breastfeeding counseling, which also statistically increased EBF practices.

**Conclusions:** The findings suggest that BEST4Baby was highly usable and accepted by mothers serving as PCs to support other mothers in their EBF practices and led to positive outcomes in the intervention group's EBF rates. The pilot study demonstrated that using the specially designed BEST4Baby app was an important support tool for mothers to serve as PCs during the 9 home visits.

**Trial Registration:** Clinicaltrials.gov NCT03533725; <https://clinicaltrials.gov/ct2/show/NCT03533725>

**KEYWORDS**

mobile health; mHealth; peer counselors; breastfeeding; rural; usability; low- and middle-income countries; agile; task shifting; user-centered design; mobile phone

## Introduction

### Exclusive Breastfeeding

Exclusive breastfeeding (EBF), defined as *the practice of giving an infant only breast milk for the first 6 months of life*, has tremendous benefits for the mother-infant dyad. It protects the infant from infections, sudden infant death syndrome, dental malocclusions, and the development of obesity and diabetes later in life [1-4]. EBF offers protection for the mother from breast cancer, ovarian cancer, and postpartum depression and improves the birth spacing interval [1,3,4]. Of great significance, EBF has been shown to be most effective in reducing overall infant mortality [5], which could prevent an estimated 823,000 child deaths [2-4] and an annual economic loss of approximately US \$302 billion. EBF practices can significantly contribute to improved outcomes in under resourced environments [6].

Low- and middle-income countries (LMICs) have the highest infant mortality rates (IMRs), and approximately 63% of infants aged <6 months are not exclusively breastfed [3]. This is especially important in India, which has reported an IMR of 30 per 1000 live births. India's IMR exceeds that of several other LMICs in Asia, such as Uzbekistan (19 per 1000 live births), Bangladesh (25 per 1000 births), and Egypt (18 per 1000 live births). Despite the Indian government's longstanding efforts to increase breastfeeding rates, EBF remains low. The National Family Health Survey 2015 to 2016 estimated that only 54.9% of infants aged <6 months are exclusively breastfed in India [7]. Furthermore, EBF rates are not homogenous in India and vary by region [8], with rates in Southern India reported to be 43.7% from infancy to the age of 5 months [8].

### Socioecological Factors of EBF

Multiple socioecological factors affect EBF rates in India. The social determinants of health include a mother's access to health care and health information, education, income, and area of residence [9], all of which affect the rates of EBF. The mother's lack of knowledge on the importance of EBF to overall health, inappropriate breastfeeding techniques, and late initiation of breastfeeding are barriers at the individual level [10-13]. At the societal level, discarding colostrum, prelacteal feeding, lack of family support, lack of the mother's autonomy in decision-making (eg, mother-in-law's influence on breastfeeding practice), and poor counseling regarding key aspects of EBF are cited as significant barriers [9,12,13]. In South Asia, programs that used repeated exposures for counseling and education on breastfeeding practices during pregnancy and in the early postpartum period were more likely to be effective in improving breastfeeding rates. Short programs that contained irregular exposures, poor timing, and inadequate coverage of the target population were less likely to affect breastfeeding rates.

### Breastfeeding Support System, the Role of Community Health Workers, and the Importance of Task-Shifting Breastfeeding Counseling In India

Task shifting is a delegation process in which tasks are moved from a highly specialized workforce to less specialized health workers, where appropriate. The Indian government has implemented and deployed different types of paid or incentivized community health workers (CHWs) as part of India's health care delivery system to task shift various health care activities [14,15]. Among the 3 cadres of CHWs implemented, auxiliary nurse midwives (ANMs) are based at a subcenter and visit villages and provide care at the subcenter. ANMs are supported by Anganwadi workers, who work solely in their villages and focus on providing food supplements to young children, adolescent girls, and lactating women. Finally, Accredited Social Health Activists (ASHAs) are the largest cadres of CHWs. They have been deployed to supplement the work of ANMs and Anganwadi workers [15]. ASHAs provide health promotion, specifically regarding nutrition, sanitation and hygiene, preparedness for birth and safe delivery, immunization, breastfeeding, complementary feeding, and prevention of common infections. Although they provide a valuable contribution to supporting and promoting maternal and child health in their communities, ASHAs often feel rushed, tired, overworked, and underpaid [16].

Studies using CHWs have generally shown significant improvements in general maternal-infant care practices (eg, skin-to-skin care) and some areas of breastfeeding support, including initiation and complementary feeding [17,18]. Furthermore, establishing a network of CHWs who can educate, support, and make necessary referrals is linked to increasing EBF [19]. Although paid or incentivized and trained CHWs have been used for breastfeeding interventions in India, the use of mothers as unpaid peer counselors (PCs) for breastfeeding has not been explored.

### Technology-Based Support for CHWs

Mobile technology for health interventions presents a strong opportunity for improving breastfeeding practices compared with usual care, including the cost-effectiveness of using CHWs such as ASHAs. Mobile health (mHealth) apps that rely on wireless access to the internet are common in India, where broad access is available, including in rural areas. Several mHealth interventions have been tested in rural areas in India to improve maternal-child health outcomes, including breastfeeding [17,18,20]. These mHealth intervention studies using CHWs [17], such as ASHAs [18], have reported improvements, including increases in job confidence [17], coordination [17], coverage [18], and quality of services in hard-to-reach areas [18]. Despite established evidence of the effectiveness of using CHWs to provide community-based counseling and education to improve EBF rates [9,21-25], the use of mHealth with task

shifting to unpaid mothers as PCs in India has yet to be explored. Brief training with mothers with prior breastfeeding experiences to serve as PCs, using mHealth technology as suitable support tools, within rural Indian communities can further expand the types of available CHWs. They would further address the need for community-based peer support to promote breastfeeding in rural India [26]. Thus, Breastfeeding Education Support Tool for Baby (BEST4Baby) is a suitable medium for training and supporting unpaid PCs in providing in-home breastfeeding counseling support to mothers in rural India.

## Methods

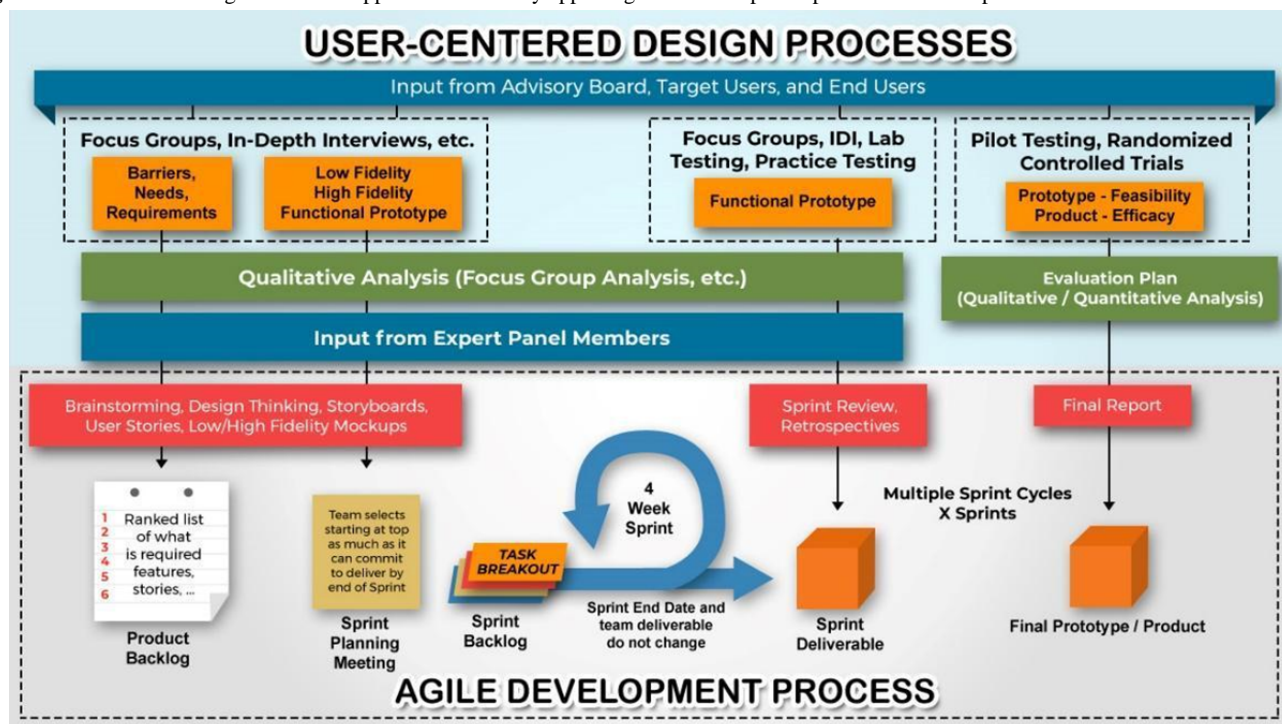
### Overview

This paper describes the design, development (ie, content and technical), and testing of the mHealth app called the *BEST4Baby* to support a community-based task-shifting intervention that uses trained mothers as PCs to provide counseling and education to mothers to achieve an improved EBF rate in the Belagavi district of Karnataka, India. The investigators included partners at Thomas Jefferson University (TJU; Philadelphia, Pennsylvania, United States) and investigators at Jawaharlal Nehru Medical College of Karnataka Lingayat Education, Academy of Higher Education and Research (Belagavi, India),

and Benten Technologies (Benten; Manassas, Virginia, United States).

The team leveraged a user-centered design (UCD) approach to design and develop the initial version of the BEST4Baby mHealth app. This approach was used to better understand the users, counseling tasks, and the environment by involving PCs and key stakeholders throughout the design process to create a positive user experience [27]. This process ensured that the BEST4Baby counseling program and the training materials, mobile app, and job-aid tools were acceptable and feasible in advance of pilot testing. The team also leveraged an agile development methodology for content and technology development. The agile development methodology emphasizes iterative development and integrates feedback from all key stakeholders in the development process to refine the BEST4Baby app over time, including the content and app [28]. All design and development processes were initially conducted in English to facilitate communication among all research team members during the study. Meetings were conducted web-based via UberConference and Zoom. Figure 1 illustrates the UCD process implemented alongside the agile development of the BEST4Baby mHealth app. The mHealth app was then pilot tested for usability with PCs to deliver effective EBF counseling to mothers at home.

**Figure 1.** The Breastfeeding Education Support Tool for Baby app design and development process. IDI: in-depth interview.



### BEST4Baby App Design

#### Overview

Formative qualitative research was conducted with a breastfeeding advisory panel and mothers with and without breastfeeding experience [19]. The members of the breastfeeding advisory panel included local clinicians and academics incorporating the medical disciplines of obstetrics and pediatrics, representatives of advocacy groups, a Karnataka State Ministry

of Health official, the Reproductive and Child Health Officer, and the District Health Officer from Belgaum, India. The results of the qualitative research and analyses helped the team to create an initial product backlog for the BEST4Baby mHealth app. Microsoft PowerPoint was used as a design tool to iteratively create mock-ups for the BEST4Baby mHealth app. The design for the initial mock-ups was based on information derived from formative qualitative research and informed the app's structure of number, timing, spacing of visits, and specific content that would be reviewed during the visits and incorporated cultural



practices during various stages of breastfeeding. Initially, the BEST4Baby app focused on addressing the importance of identifying salient barriers to breastfeeding.

### **Design Iterations**

The design versions for the BEST4Baby app included the (1) initial mock-up design, (2) revised mock-up design with integrated educational content, and (3) final interactive wireframes with content. The stakeholders were engaged in the feedback of the design mock-ups to ensure proper flow and integration with BEST4Baby PC training. Changes in the design included simplifying navigation options, removing one prenatal visit, and adding the potential to make unscheduled visits. The final design integrated all the educational content and was reviewed and approved by our research collaborators in India.

### **Visit Session Design**

The BEST4Baby app design centered on the following concepts: (1) time (relatively short sessions), (2) content (small, bite-size units of information to be covered in each session), (3) curriculum (personalized content for each session as part of a curriculum), (4) form (all sessions presented in the same, consistent structure and format, as well as incorporating multimedia when appropriate), and (5) flexibility (different times and locations for each delivery) [29]. Mothers who served as PCs were the primary users of the app. After brief training, the PCs could use the mHealth app to master complex breastfeeding counseling topics by using many short, guided visit sessions with a built-in refresher training module for use in the field. Each visit session was designed to dynamically present content for PCs to share and counsel mothers based on their responses at each visit. The app was designed to provide just-in-time information without overloading mothers with too much information. With scheduled visits by PCs at various stages (before and after delivery), mothers were only presented with stage-appropriate content for that period. For example, visit 2 was 32 to 36 weeks ante partum and provided information on the importance of colostrum; visit 3, which occurs within 1 to 3 days after delivery, focuses on issues related to delivery, such as baby delivery questions (baby's weight, birthing hospital, and delivery method), feeding questions, assessment of baby and mother's health, and others. The app was also designed with assessments for PCs conducted at the beginning of each visit to identify potential problems. On the basis of the assessment results, PC counseling content was dynamically shown to address specific solutions for mothers.

## **BEST4Baby Content Development**

### **Overview**

In parallel with the app design, the educational content for the BEST4Baby mHealth intervention was developed with experts in the field. The educational content leveraged the formative research with mothers. It noted content that was successful and unsuccessful in promoting EBF [19], results from a prior study, and input from PCs to finalize the design. The educational content was based on a modified, culturally adapted World Health Organization breastfeeding counseling course. The content was created for training PCs and, as appropriate, for use during in-home and hospital visits to improve the mother's

knowledge regarding EBF and optimal feeding practices [30]. The content included videos comprising brief but common experiences prevalent in breastfeeding practices. The video content was tested for compatibility with Android and iOS devices.

### **Behavioral Theories**

The design of the BEST4Baby educational content leveraged the Social Cognitive Theory (SCT) and the Theory of Planned Behavior (TPB) [31]. The SCT's construct of *self-efficacy* was used to develop educational content using videos of other Indian mothers who were successful in various breastfeeding techniques. The SCT's constructs of *outcomes expectancy* and *observational learning* were incorporated into design components that addressed sociostructural factors that facilitate or hinder pursued behavior, such as the mother or mother-in-law's influence on breastfeeding, prelacteal feeding, and complementary feeding. The TPB's construct of *behavioral intentions* was incorporated into the BEST4Baby app design. The TPB constructs of *attitude* (eg, assessing initial and subsequent attitudes), *subjective norms* (eg, involving key individuals such as mothers-in-law), and *perceived behavioral controls* (eg, reinforcing a mother's self-efficacy using an observational assessment, personalized counseling to their breastfeeding challenges, and videos to perform the behavior) were incorporated into the step-by-step counseling guide [31,32].

### **Content Integration With App**

The educational content was integrated into the BEST4Baby app design as each component of the content was completed. After incorporating this design and content into an interactive wireframe, a focus group session was conducted with 7 PCs to demonstrate the mHealth app and obtain feedback. After the focus group, we administered the System Usability Scale (SUS) survey to evaluate the initial usability of the BEST4Baby app design, obtain final feedback on the design, and incorporate feedback before completing the technological development of the prototype. The SUS is a reliable scale widely used and validated to measure system usability, independent of the type of technology [33]. SUS has also been used in various languages and cultures, including in India [34,35]. It has become an industry standard, with reference in >600 publications [36].

## **BEST4Baby Technological Development**

The Benten development team consulted TJU off site and the Indian team as product owners during the agile development. In addition, lessons learned from a prior mHealth intervention, called Community Level Interventions for pre-eclampsia [37] in India, provided knowledge on potential infrastructure challenges for the use of mobile technology in rural India.

### **Overall Architecture**

The mHealth app was developed with cross-platform mobile technologies using MongoDB, Express, ReactJS, and NodeJS, which is a technology stack available for any Android or iOS device. The initial architecture included web-based and offline capabilities to address wireless coverage issues. The team also developed security features, such as strong password protection for authentication, access control, and secure transport of information via the Secure Sockets Layer, to ensure that the

mHealth app conformed to India's regulations on the privacy and security of patient data. The prototype was also designed to capture data regarding app use to assist in the evaluation of the app. The content leveraged XML to encode the content and allowed for a dynamic, personalized presentation of breastfeeding counseling information based on the data captured for each mother during their visit. At the initial visit, detailed data of app use were captured, including GPS data, time spent by PCs in each content area, and individual responses from mothers during each counseling session.

### ***App Features***

The team created the BEST4Baby app to provide a systematic step-by-step guide [26,27] to control the quality of PC counseling at each visit and promote optimal breastfeeding practices in mothers. BEST4Baby's unique features included (1) a step-by-step guide for each visit on proper breastfeeding counseling and education, (2) systematic breastfeeding assessments for mothers during the initial prenatal and postpartum period to shape beliefs and troubleshoot breastfeeding challenges, (3) time-sequenced prenatal and

postpartum period education to deliver appropriate just-in-time breastfeeding information, and (4) personalized communication to address specific breastfeeding challenges of lactating mothers during the postpartum period. The PCs accessed the BEST4Baby app using a simplified log-in process. Once authenticated, a PC could view the educational content and conduct in-home visits in English or the local language (Figure 2). The BEST4Baby home page provides access to functionality, such as the (1) Mom List, which allows PCs to add to and track visits for each designated mother, including completed and pending visits; (2) Support List, which allows PCs to communicate with clinicians and technical support as needed; (3) Visit Content, which provides step-by-step guided counseling incorporated within the educational content, including multimedia and video clips, organized into 9 sessions corresponding to the visits by PCs to mothers (2 antenatal and 7 post partum); (4) Appointments, which allows PCs to view upcoming visits, schedule new appointments, and add them to the calendar; (5) Video, which provides PCs with a breastfeeding content library for viewing directly whenever needed during the visit.

**Figure 2.** Screenshots of the Breastfeeding Education Support Tool for Baby mobile health app. A. Login screen of the BEST4Baby app B. BEST4Baby Homepage C. App displaying visit numbers and assigned moms D. App displaying support list for technical and clinical support E. App displaying visit content (In English) F. App displaying visit content (In Kannada) G. App displaying appointment calendar H. App displaying educational videos on breastfeeding.



**Integration and Testing Preparation**

After development, the Indian team translated the final content into the local language, Kannada. The translation was provided in Unicode for incorporation into the BEST4Baby app. The app was designed to display the content and interface components in English and Kannada. In preparation for the pilot study, the

Indian team researched possible low-cost Android devices, which constitutes >90% of the market in India [38]. Different Android tablet devices were considered. A Samsung Android tablet was chosen as it was an affordable, common, easily accessible, and portable device that could fit in the BEST4Baby kit for PCs to carry to the pilot testing site. The kit had the following items: (1) a life-size newborn doll to demonstrate

positioning, (2) a skin-colored sock to prepare a breast model for demonstrating proper latching, (3) a digital scale for weighing and assessing the growth of the infant, (4) a nipple plunger to mitigate the problem of an inverted nipple, and (5) a Samsung Android tablet with wireless and GPS capability preloaded with the BEST4Baby app and secured to allow for its sole use with the app. A total of 25 Samsung tablet devices were purchased for use by the PCs for the study implementation. Final testing was performed to ensure compatibility of the final BEST4Baby software with the device selected for the PCs.

## Usability Testing

### Study Participants

A total of 56 potential PCs were identified by staff from 5 local primary health centers in the Belagavi health district, facilitated through announcements and word of mouth [26]. Out of 56 PCs, 25 (45%) were selected for usability testing of the platform based on the inclusion criteria of (1) residing in the local community; (2) having breastfed within the past 5 years; (3) having at least 10 years of formal education; (4) having an available mobile phone; (5) being familiar with operating an Android phone; and (6) being able to read, write, and communicate in the local language.

### Study Design

Usability refers to “the quality of a user’s experience when interacting with products or systems, including websites, software, devices, or applications” [39]. A sample size of 20 PCs was sufficient to achieve 80% power to detect a difference of 6.0 between the actual mean of 74.0 and the null-hypothesized mean of 68.0. The International Organization of Standardization 9241-11 considers usability to be a measure of the system’s technical effectiveness, efficiency, and satisfaction from the perspective of user experiences [40]. BEST4Baby usability was tested in a real-life setting within 6 clusters of the Global Research Network area located in Belagavi district (Karnataka, India). The pilot study involved pretesting for usability with 25 PCs and posttesting for usability with 22 PCs (88%; n=3, 12% dropped out from the study) recruited from the study areas.

### Study Procedure

Following recruitment, all 25 PCs attended a 3-day training session, which included content on breastfeeding knowledge and skills, counseling techniques, and instruction on the BEST4Baby mHealth app. Each PC received an ID badge and a branded BEST4Baby mobile device containing the BEST4Baby mobile app. During the posttraining time point, each PC was assigned to 5 mothers in the intervention group, who resided in their community. PCs used the BEST4Baby app during home visits to mothers following a 9-visit schedule at >28 to 32 weeks ante partum, 32 to 36 weeks ante partum, postpartum days 1 to 3, postpartum day 7, postpartum day 15, postpartum 1 month, postpartum 2 months postpartum 4 months, and postpartum 6 months. As the PCs went to the visit sessions with the mothers, they manually entered and collected data using the BEST4Baby app on the assigned Samsung tablet. Data from the Samsung tablet automatically synchronizes to a server if the internet is available. The BEST4Baby app was also

designed with the capability of an offline mode feature, which means that the educational content for the visits was stored on the Samsung tablet, and data could be collected and stored on the tablet without the need to access the internet or wireless service to address potential wireless coverage issues in a rural area. The offline mode allowed the data to be synchronized later to the server whenever the app could detect internet service availability. In addition, the app also had a feature to enable PCs to synchronize the data manually. At the end of the training, 25 PCs were asked to complete a SUS survey to assess the usability and acceptability of the BEST4Baby app. The SUS questionnaire was translated into the local language. The results from the SUS score were found to be highly usable, with an average score of 87.5 (SD 8.2; range 72.5-100).

### Instruments

BEST4Baby’s usability was assessed using the SUS [33,36], a widely used validated scale that can be used with a variety of technologies and provides a single usability score [36]. The survey comprises a 10-item Likert scale for respondents (with 5-point anchors from strongly agree to strongly disagree). Scores for each item in the SUS survey are converted to a number, added together, and then multiplied by 2.5 to create a single SUS score between 0 and 100. A SUS score >68 is considered above average and supports acceptability for use [36]. Data regarding app use were automatically captured from the app and obtained from the server to assess user engagement with the BEST4Baby app.

### Usability Data Analysis

Data were collected from the surveys, deidentified, and entered into a Microsoft Excel spreadsheet. After verifying the data, baseline demographics were assessed using descriptive statistics. Individual SUS scores were calculated for each participant, and a mean SUS score with SD was provided for postpilot survey results. App use data were obtained from the server after the pilot and analyzed using descriptive statistics.

### Ethics Approval

Institutional review board approval was obtained from Karnataka Lingayat Education University (IRB registration number: 00008025), Jawaharlal Nehru Medical College in India, and TJU in the United States.

## Results

### Participant Characteristics

A total of 25 participants were recruited as PCs, of whom 3 (12%) did not complete the study, resulting in 22 (88%) participants who completed both the study and posttesting usability survey. The usability survey was translated into the local language before it was administered. The participants were all female, with a mean age of 30.18 (SD 4.43) years [26]. Half of the participants reported having at least 8 to 10 years of education, with the remainder having 11 to 16 years of education; all had prior experience of using a smartphone for an average of 14.95 months (Table 1).

**Table 1.** Demographics (N=22).

Characteristics	Values
Age (years), mean (SD)	30.18 (4.43)
<b>Age (years), n (%)</b>	
21-25	3 (14)
26-30	11 (50)
31-35	6 (27)
36-40	2 (9)
<b>Years of education, n (%)</b>	
8-10	11 (50)
11-16	11 (50)
Smartphone experience (months), mean (SD)	14.95 (21.33)

### App Use Data

A total of 22 PCs completed the 9-visit schedule with mothers, with each PC serving 4 to 5 mothers. PCs also had the option of adding additional unscheduled visits (a feature that was optional and accessible in the app). Unscheduled visits were designed for the PCs to provide additional breastfeeding support if needed. The time that the PCs spent on the app with the mothers during each visit ranged from 6.6 to 39.6 minutes. The longest period was during the first visit, whereas the shortest was during the unscheduled visits. The average time spent by PCs on visits 1 to 5 was 29.1 minutes, whereas the average time for later visits (eg, visits 6-9) was 12.7 minutes (Table 2). After the first 6 visits, PCs spent an average of <12 minutes on the last 3 visits (2-, 4-, and 6-month visits post partum).

Unscheduled visits (ie, visits that were not originally planned) also had the shortest duration (<7 minutes).

During each visit, PCs could revisit the educational content with mothers, which was provided previously (another optional feature of the app) based on individual needs. The most frequently revisited educational content topics included the importance of colostrum feeding, position and attachment, and proper burping after feeding. The content related to the importance of colostrum feeding, the many advantages of breastfeeding, and the dangers of artificial and prelacteal feeding were also revisited during visit 3 (Table 3). Content related to the demonstration of position and attachment, expression of breast milk, increased secretion of breast milk, and proper burping after a feed was revisited during visits 4 and 5 (Table 3).

**Table 2.** Visit duration.

Visit name	Targeted visit time points	Average duration per visit	Duration per screen, mean (SD)
Visit 1	28-32 weeks ante partum	39.6	1.41 (1.69)
Visit 2	32-36 weeks ante partum	24.0	1.09 (1.50)
Visit 3	Postpartum days 1-3	28.1	0.76 (0.82)
Visit 4	Postpartum day 7	28.5	0.81 (0.95)
Visit 5	Postpartum day 15	25.4	0.73 (0.89)
Visit 6	Postpartum 1 month	16.6	0.57 (0.50)
Visit 7	Postpartum 2 months	11.2	0.56 (0.37)
Visit 8	Postpartum 4 months	11.0	0.55 (0.40)
Visit 9	Postpartum 6 months	11.9	0.56 (0.40)
Unscheduled visits	As needed between visits	6.6	0.28 (0.37)

**Table 3.** Topics revisited and time frame of when it occurred.

Content topic	Revisits, N	When it occurred
Flat or inverted nipples	3	Unscheduled visits
Breast engorgement	4	Unscheduled visits
Sore nipples	2	Unscheduled visits
Importance of colostrum feeding	95	Visit 3
Advantages of breastfeeding	32	Visit 3
Dangers of artificial feeding	23	Visit 3
Dangers of prelacteal feeding	24	Visit 3
Demonstrate position and attachment	111	Visits 4 and 5
Expression of breast milk	66	Visits 4 and 5
How to increase secretion of breast milk	58	Visits 4 and 5
Proper burping after a feed	116	Visits 4 and 5

### Usability Analysis

The SUS scores for PCs ranged from 60 to 97.5, with a mean SUS score of 85.3 (SD 9.1). The mean SUS score, which was above the *acceptable* score [33], was considered above average by industry standards, indicating high usability. The average SUS score achieved by PCs places the BEST4Baby mHealth app above the 95th percentile in terms of usability [41]. All 22 participants strongly agreed with the following statements: “I

would use this app frequently,” “I found the app easy to use,” and “It was easy to use the app during training.” In addition, 91% (20/22) of PCs strongly agreed with the statements that “The app features are well integrated” and “I can easily learn to use this app.” Most PCs strongly disagreed with the statements that “There was inconsistency in the app” and “I found the app too complex.” However, half of the PCs strongly agreed with “I needed to learn many things before I could use the app” (Table 4).

**Table 4.** Individual SUS<sup>a</sup> item scores (N=22).

SUS item	Statement (rank your impression from 1 to 5; 1=strongly disagree and 5=strongly agree)	Values, mean (SD)	5=strongly agree, n (%)	1=strongly disagree, n (%)
1	I would use this app frequently	5 (0)	22 (100)	0 (0)
2	I found the app too complex	1.6 (0.8)	0 (0)	13 (59)
3	I found the app easy to use	5 (0)	22 (100)	0 (0)
4	I need a technical person to use this app	2.3 (0.9)	0 (0)	4 (18)
5	The app features are well integrated	4.9 (0.3)	20 (91)	0 (0)
6	There was inconsistency in the app	1.4 (0.9)	1 (5)	17 (77)
7	I can easily learn to use this app	4.6 (1.2)	20 (91)	2 (9)
8	I found the app cumbersome to use	1.5 (1.1)	1 (5)	16 (73)
9	It was easy to use the app during training	5 (0)	22 (100)	0 (0)
10	I needed to learn many things before I could use the app	3.6 (1.6)	11 (50)	4 (18)

<sup>a</sup>SUS: System Usability Scale.

## Discussion

### Principal Findings

We piloted BEST4Baby with 22 PCs to design, develop, and test the usability of this mHealth app, specifically designed to support PCs in providing in-home breastfeeding counseling to mothers in rural areas. The results of this usability test suggest that BEST4Baby was highly usable and acceptable by PCs in supporting mothers on optimal breastfeeding practices. The BEST4Baby mHealth app was developed following industry best practices and standards for UCD and agile development

for use by PCs to support EBF practices in India. Development used a UCD approach to better understand the user experience [27]. The UCD process led to the creation of a unique app design that yielded a high usability rating by the mothers as PCs, resulting in high EBF rates at 6 months.

Based on the findings, the research team developed unique content and app features to support PCs in delivering in-home breastfeeding counseling, including the design of *just-in-time* information and skill-building content for mothers just before they require the information, as well as the use of checklists to guide the assessments, which included dynamic content delivery

for each session specific to the mother's challenges or needs. With only 3 days of training on breastfeeding counseling, including the use of the BEST4Baby app, mothers serving as PCs were able to learn how to use the app to conduct visit sessions.

The BEST4Baby app was pilot tested in ecologically valid settings in the Belagavi district in India and was determined to be highly usable and acceptable by PCs in this study. PCs used the BEST4Baby mHealth app to support a 9-visit intervention. The app's features, including calendar scheduling, reminders, and dynamically generated content, helped support mothers to serve as PCs to meet breastfeeding counseling visit schedules designed for the intervention. Initial visits lasted longer, whereas later visits were shorter (eg, an average of 29 minutes vs an average of <12 minutes). The most frequently revisited educational topics were related to the importance of colostrum feeding, position and attachment, and proper burping after feeding. In addition, the results related to content topics and when they were discussed were based on the specific needs to address each mother's challenge during the visits. The findings from the PCs will help inform future modifications of the BEST4Baby software.

A unique aspect of our mHealth app was the step-by-step guide with breastfeeding assessment, which provided dynamic counseling content to support mothers and created an intuitive flow for each visit. A step-by-step guide is a crucial component for activating desired behavior, as applied in evidence-based digital health interventions for mental health and behavior change [42]. Using a step-by-step guide along with a calendar for the scheduling of visits, appointment reminders, breastfeeding assessments, provision of dynamic counseling content, and use of the training module for field practice and refreshers resulted in PCs reporting high usability in supporting mothers to exclusively breastfeed.

Another unique feature of the tested mobile app was the tracking system of the PCs' adherence to the 9-visit protocol. User engagement can be a challenge when attempting to implement an mHealth intervention [43]. Future research could incorporate components such as gamification to ensure user adherence to the mobile app [44] and protocol content.

### Technical Challenges

During pilot testing, the research team experienced several technical challenges. The first occurred during the deployment of the BEST4Baby app, including data synchronization with the cloud-based server. This was resolved by providing a manual synchronization button. Second, owing to synchronization issues

across multiple devices, the research team and PCs could not log into different devices, preventing researchers from monitoring and tracking the progress of each PC in real-time. The team developed a work-around monitoring functionality by creating a weekly export report in Microsoft Excel to overcome this issue. The report provided weekly details so that each PC's visit history, including completed visits for each mother, could be provided to the research team.

### Strengths

Studies have shown high levels of perceived acceptability of mHealth-supported interventions among CHWs in low-resource settings and LMICs, including in India [45]. A recent meta-analysis of studies conducted in 6 countries suggested that mHealth may be associated with improved maternal breastfeeding attitudes, knowledge, initiation, and EBF duration [46]. However, to the best of our knowledge, BEST4Baby is among the first mHealth apps to test the use of mHealth among unpaid PCs with a limited amount of training in rural India to successfully promote and support optimal breastfeeding practices. To date, no assessment has been made on the use of mHealth to enhance breastfeeding PC programs in India.

### Limitations

This study had certain limitations. First, only feedback from mothers who served as PCs was included in the initial design process, and the input of mothers served by the PCs was not sought. Second, usability and acceptability were collected at the beginning and end of the pilot for the PCs who had breastfeeding experiences within 5 years; however, the perspective on the usability of BEST4Baby from current breastfeeding mothers was not collected. Third, owing to our small sample size of PCs to evaluate feasibility (usability), the small sample size was not powered to evaluate the study's external validity. Finally, we were also unable to explore the characteristics of PCs who reported below-average SUS scores because of our small sample size.

### Conclusions

Our findings suggest that an mHealth tool such as the BEST4Baby app can effectively help train PCs in supporting and counseling mothers in rural India to exclusively breastfeed. This study contributes to the growing literature demonstrating the applicability of a UCD with an iterative agile development for creating an mHealth app that is most usable for mothers to serve as breastfeeding PCs. The BEST4Baby app was found to be easy to use in support of breastfeeding efforts and provides a framework for its use in future trials.

---

### Acknowledgments

This study was funded by the National Institute of Health's Fogarty International Center R21 study. The sponsor had no role in the study design, data collection, data analysis, data interpretation, or writing of the report or manuscript. The corresponding author (TM) was responsible for the decision to publish.

---

### Conflicts of Interest

None declared.

## References

1. Slusser W. Breastfeeding and maternal and infant health outcomes in developed countries. *AAP Grand Rounds* 2007 Aug 01;18(2):15-16. [doi: [10.1542/gr.18-2-15](https://doi.org/10.1542/gr.18-2-15)]
2. Sankar MJ, Sinha B, Chowdhury R, Bhandari N, Taneja S, Martinez J, et al. Optimal breastfeeding practices and infant and child mortality: a systematic review and meta-analysis. *Acta Paediatr* 2015 Dec 04;104(467):3-13. [doi: [10.1111/apa.13147](https://doi.org/10.1111/apa.13147)] [Medline: [26249674](https://pubmed.ncbi.nlm.nih.gov/26249674/)]
3. Victora CG, Bahl R, Barros AJ, França GV, Horton S, Krasevec J, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *Lancet* 2016 Jan;387(10017):475-490. [doi: [10.1016/S0140-6736\(15\)01024-7](https://doi.org/10.1016/S0140-6736(15)01024-7)]
4. Smith ER, Hurt L, Chowdhury R, Sinha B, Fawzi W, Edmond KM, Neovita Study Group. Delayed breastfeeding initiation and infant survival: a systematic review and meta-analysis. *PLoS One* 2017;12(7):e0180722 [FREE Full text] [doi: [10.1371/journal.pone.0180722](https://doi.org/10.1371/journal.pone.0180722)] [Medline: [28746353](https://pubmed.ncbi.nlm.nih.gov/28746353/)]
5. Breastfeeding policy. World Health Organization. URL: [http://www.who.int/nutrition/publications/globaltargets2025\\_policybrief\\_breastfeeding/en/](http://www.who.int/nutrition/publications/globaltargets2025_policybrief_breastfeeding/en/) [accessed 2021-08-04]
6. Rollins NC, Bhandari N, Hajeebhoy N, Horton S, Lutter CK, Martinez JC, et al. Why invest, and what it will take to improve breastfeeding practices? *Lancet* 2016 Jan;387(10017):491-504. [doi: [10.1016/s0140-6736\(15\)01044-2](https://doi.org/10.1016/s0140-6736(15)01044-2)]
7. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS; 2017.
8. Ogbo FA, Dhami MV, Awosemo AO, Olusanya BO, Olusanya J, Osuagwu UL, et al. Regional prevalence and determinants of exclusive breastfeeding in India. *Int Breastfeed J* 2019 May 16;14(1):20. [doi: [10.1186/s13006-019-0214-0](https://doi.org/10.1186/s13006-019-0214-0)]
9. Sharma IK, Byrne A. Early initiation of breastfeeding: a systematic literature review of factors and barriers in South Asia. *Int Breastfeed J* 2016 Jun 18;11(1):17 [FREE Full text] [doi: [10.1186/s13006-016-0076-7](https://doi.org/10.1186/s13006-016-0076-7)] [Medline: [27330542](https://pubmed.ncbi.nlm.nih.gov/27330542/)]
10. Garg R, Deepti S, Padda A, Singh T. Breastfeeding knowledge and practices among rural women of Punjab, India: a community-based study. *Breastfeed Med* 2010 Dec;5(6):303-307. [doi: [10.1089/bfm.2010.0005](https://doi.org/10.1089/bfm.2010.0005)] [Medline: [20925495](https://pubmed.ncbi.nlm.nih.gov/20925495/)]
11. Majra JP. Barriers to early initiation and continuation of breastfeeding in a tertiary care institute of Haryana: a qualitative study in nursing care providers. *J Clin Diagnos Res* 2016;10(9). [doi: [10.7860/jcdr/2016/19072.8559](https://doi.org/10.7860/jcdr/2016/19072.8559)]
12. Bhandari D, Pandya Y, Sharma D. Barriers to exclusive breastfeeding in rural community of central Gujarat, India. *J Family Med Prim Care* 2019 Jan;8(1):54-61 [FREE Full text] [doi: [10.4103/jfmpc.jfmpc\\_329\\_18](https://doi.org/10.4103/jfmpc.jfmpc_329_18)] [Medline: [30911481](https://pubmed.ncbi.nlm.nih.gov/30911481/)]
13. Chugh Sachdeva R, Mondkar J, Shanbhag S, Manuhar M, Khan A, Dasgupta R, et al. A qualitative analysis of the barriers and facilitators for breastfeeding and Kangaroo mother care among service providers, mothers and influencers of neonates admitted in two urban hospitals in India. *Breastfeed Med* 2019 Mar;14(2):108-114. [doi: [10.1089/bfm.2018.0177](https://doi.org/10.1089/bfm.2018.0177)] [Medline: [30676061](https://pubmed.ncbi.nlm.nih.gov/30676061/)]
14. Koehn HJ, Zheng S, Houser RF, O'Hara C, Rogers BL. Remuneration systems of community health workers in India and promoted maternal health outcomes: a cross-sectional study. *BMC Health Serv Res* 2020 Jan 20;20(1):48 [FREE Full text] [doi: [10.1186/s12913-019-4883-6](https://doi.org/10.1186/s12913-019-4883-6)] [Medline: [31959157](https://pubmed.ncbi.nlm.nih.gov/31959157/)]
15. India's auxiliary nurse-midwife, anganwadi worker, accredited social health activist, multipurpose worker, and lady health visitor programs. CHW Central. URL: <https://chwcentral.org/indias-auxiliary-nurse-midwife-anganwadi-worker-accredited-social-health-activist-multipurpose-worker-and-lady-health-visitor-programs/> [accessed 2022-04-08]
16. Kawade A, Gore M, Lele P, Chavan U, Pinnock H, Smith P, RESPIRE collaboration. Interplaying role of healthcare activist and homemaker: a mixed-methods exploration of the workload of community health workers (Accredited Social Health Activists) in India. *Hum Resour Health* 2021 Jan 06;19(1):7 [FREE Full text] [doi: [10.1186/s12960-020-00546-z](https://doi.org/10.1186/s12960-020-00546-z)] [Medline: [33407518](https://pubmed.ncbi.nlm.nih.gov/33407518/)]
17. Carmichael S, Mehta K, Srikantiah S, Mahapatra T, Chaudhuri I, Balakrishnan R, Ananya Study Group\*. Use of mobile technology by frontline health workers to promote reproductive, maternal, newborn and child health and nutrition: a cluster randomized controlled Trial in Bihar, India. *J Glob Health* 2019 Dec;9(2):0204249. [doi: [10.7189/jogh.09.020424](https://doi.org/10.7189/jogh.09.020424)] [Medline: [31788233](https://pubmed.ncbi.nlm.nih.gov/31788233/)]
18. Modi D, Dholakia N, Gopalan R, Venkatraman S, Dave K, Shah S, et al. mHealth intervention "ImTeCHO" to improve delivery of maternal, neonatal, and child care services-A cluster-randomized trial in tribal areas of Gujarat, India. *PLoS Med* 2019 Oct 24;16(10):e1002939 [FREE Full text] [doi: [10.1371/journal.pmed.1002939](https://doi.org/10.1371/journal.pmed.1002939)] [Medline: [31647821](https://pubmed.ncbi.nlm.nih.gov/31647821/)]
19. Charantimath U, Bellad R, Majantashetti N, Washio Y, Derman R, Kelly PJ, et al. Facilitators and challenges to exclusive breastfeeding in Belagavi District, Karnataka, India. *PLoS One* 2020 May 4;15(5):e0231755 [FREE Full text] [doi: [10.1371/journal.pone.0231755](https://doi.org/10.1371/journal.pone.0231755)] [Medline: [32365108](https://pubmed.ncbi.nlm.nih.gov/32365108/)]
20. Balakrishnan R, Gopichandran V, Chaturvedi S, Chatterjee R, Mahapatra T, Chaudhuri I. Continuum of Care Services for Maternal and Child Health using mobile technology - a health system strengthening strategy in low and middle income countries. *BMC Med Inform Decis Mak* 2016 Jul 07;16(1):84 [FREE Full text] [doi: [10.1186/s12911-016-0326-z](https://doi.org/10.1186/s12911-016-0326-z)] [Medline: [27387548](https://pubmed.ncbi.nlm.nih.gov/27387548/)]
21. Kushwaha KP, Sankar J, Sankar MJ, Gupta A, Dadhich JP, Gupta YP, et al. Effect of peer counselling by mother support groups on infant and young child feeding practices: the Lalitpur experience. *PLoS One* 2014 Nov 4;9(11):e109181 [FREE Full text] [doi: [10.1371/journal.pone.0109181](https://doi.org/10.1371/journal.pone.0109181)] [Medline: [25369452](https://pubmed.ncbi.nlm.nih.gov/25369452/)]



22. Daniels K, Sanders D, Daviaud E, Doherty T. Valuing and sustaining (or not) the ability of volunteer community health workers to deliver integrated community case management in northern Ghana: a qualitative study. *PLoS One* 2015 Jun 16;10(6):e0126322 [FREE Full text] [doi: [10.1371/journal.pone.0126322](https://doi.org/10.1371/journal.pone.0126322)] [Medline: [26079713](https://pubmed.ncbi.nlm.nih.gov/26079713/)]
23. Glenton C, Colvin C, Carlsen B, Swartz A, Lewin S, Noyes J, et al. Barriers and facilitators to the implementation of lay health worker programmes to improve access to maternal and child health: qualitative evidence synthesis. *Cochrane Database Syst Rev* 2013 Oct 08;2013(10):CD010414 [FREE Full text] [doi: [10.1002/14651858.CD010414.pub2](https://doi.org/10.1002/14651858.CD010414.pub2)] [Medline: [24101553](https://pubmed.ncbi.nlm.nih.gov/24101553/)]
24. Sudfeld CR, Fawzi WW, Lahariya C. Peer support and exclusive breastfeeding duration in low and middle-income countries: a systematic review and meta-analysis. *PLoS One* 2012 Sep 18;7(9):e45143. [doi: [10.1371/journal.pone.0045143](https://doi.org/10.1371/journal.pone.0045143)] [Medline: [23028810](https://pubmed.ncbi.nlm.nih.gov/23028810/)]
25. Jolly K, Ingram L, Khan KS, Deeks JJ, Freemantle N, MacArthur C. Systematic review of peer support for breastfeeding continuation: meta-regression analysis of the effect of setting, intensity, and timing. *BMJ* 2012 Jan 25;344(jan25 4):d8287. [doi: [10.1136/bmj.d8287](https://doi.org/10.1136/bmj.d8287)] [Medline: [22277543](https://pubmed.ncbi.nlm.nih.gov/22277543/)]
26. Short VL, Bellad RM, Kelly PJ, Washio Y, Ma T, Chang K, et al. Feasibility, acceptability, and preliminary impact of an mHealth supported breastfeeding peer counselor intervention in rural India. *Int J Gynaecol Obstet* 2022 Jan 20;156(1):48-54. [doi: [10.1002/ijgo.13599](https://doi.org/10.1002/ijgo.13599)] [Medline: [33454986](https://pubmed.ncbi.nlm.nih.gov/33454986/)]
27. Kruk ME, Gage AD, Arsenaault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. *Lancet Global Health* 2018 Nov;6(11):e1196-e1252. [doi: [10.1016/s2214-109x\(18\)30386-3](https://doi.org/10.1016/s2214-109x(18)30386-3)]
28. Raghu A, Praveen D, Peiris D, Tarassenko L, Clifford G. Engineering a mobile health tool for resource-poor settings to assess and manage cardiovascular disease risk: SMARThealth study. *BMC Med Inform Decis Mak* 2015 Apr 29;15(1):36 [FREE Full text] [doi: [10.1186/s12911-015-0148-4](https://doi.org/10.1186/s12911-015-0148-4)] [Medline: [25924825](https://pubmed.ncbi.nlm.nih.gov/25924825/)]
29. Simons LP, Foerster F, Bruck PA, Motiwalla L, Jonker CM. Microlearning mApp raises health competence: hybrid service design. *Health Technol (Berl)* 2015;5(1):35-43 [FREE Full text] [doi: [10.1007/s12553-015-0095-1](https://doi.org/10.1007/s12553-015-0095-1)] [Medline: [26097799](https://pubmed.ncbi.nlm.nih.gov/26097799/)]
30. Adhisivam B, Vishnu Bhat B, Poorna R, Thulasingham M, Pournami F, Joy R. Postnatal counseling on exclusive breastfeeding using video - experience from a tertiary care teaching hospital, south India. *J Matern Fetal Neonatal Med* 2017 Apr 26;30(7):834-838. [doi: [10.1080/14767058.2016.1188379](https://doi.org/10.1080/14767058.2016.1188379)] [Medline: [27160763](https://pubmed.ncbi.nlm.nih.gov/27160763/)]
31. Jamei F, Ostovar A, Javadzade H. Predictors of exclusive breastfeeding among nulliparous Iranian mothers: application of the theory of planned behavior. *Int J Pediatr* 2017;5(3):4457-4467. [doi: [10.22038/IJP.2016.20815.1748](https://doi.org/10.22038/IJP.2016.20815.1748)]
32. Bajoulvand R, González-Jiménez E, Imani-Nasab M, Ebrahimzadeh F. Predicting exclusive breastfeeding among Iranian mothers: application of the theory of planned behavior using structural equation modeling. *Iranian J Nursing Midwifery Res* 2019;24(5):323. [doi: [10.4103/ijnmr.ijnmr\\_164\\_18](https://doi.org/10.4103/ijnmr.ijnmr_164_18)]
33. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Human Comput Interact* 2008 Jul 30;24(6):574-594. [doi: [10.1080/10447310802205776](https://doi.org/10.1080/10447310802205776)]
34. Pal D, Vanijja V. Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. *Child Youth Serv Rev* 2020 Dec;119:105535 [FREE Full text] [doi: [10.1016/j.childyouth.2020.105535](https://doi.org/10.1016/j.childyouth.2020.105535)] [Medline: [33020676](https://pubmed.ncbi.nlm.nih.gov/33020676/)]
35. Gao M, Kortum P, Oswald FL. Multi-language toolkit for the system usability scale. *Int J Human Comput Interact* 2020 Aug 19;36(20):1883-1901. [doi: [10.1080/10447318.2020.1801173](https://doi.org/10.1080/10447318.2020.1801173)]
36. Measuring usability with the System Usability Scale (SUS). *MeasuringU*. URL: <https://measuringu.com/sus/> [accessed 2018-08-23]
37. Bellad MB, Vidler M, Honnungar NV, Mallapur A, Ramadurg U, Charanthimath U, CLIP Working Group. Maternal and newborn health in Karnataka state, India: the community level interventions for pre-eclampsia (CLIP) trial's baseline study results. *PLoS One* 2017 Jan 20;12(1):e0166623 [FREE Full text] [doi: [10.1371/journal.pone.0166623](https://doi.org/10.1371/journal.pone.0166623)] [Medline: [28107350](https://pubmed.ncbi.nlm.nih.gov/28107350/)]
38. Android v iOS market share 2019. *Device Atlas*. URL: <https://deviceatlas.com/blog/android-v-ios-market-share#colombia> [accessed 2022-03-04]
39. Usability evaluation basics. *usability.gov*. URL: <https://www.usability.gov/what-and-why/usability-evaluation.html> [accessed 2021-01-27]
40. British Standards Document BS EN ISO 9241-20 Ergonomics of human-system interaction. An ergonomic approach to accessibility within the ISO 9241 series. BSI. URL: <https://landingpage.bsigroup.com/LandingPage/Undated?UPI=00000000030209911> [accessed 2021-08-04]
41. Lewis J, Sauro J. Item benchmarks for the system usability scale. *J Usability Stud* 2018;13(3):158-167.
42. DeSouza J, Ovaysikia S, Pynn L. Correlating behavioral responses to fMRI signals from human prefrontal cortex: examining cognitive processes using task analysis. *J Vis Exp* 2012 Jun 20(64):3237 [FREE Full text] [doi: [10.3791/3237](https://doi.org/10.3791/3237)] [Medline: [22759999](https://pubmed.ncbi.nlm.nih.gov/22759999/)]
43. Wickliffe J, Kelly PJ, Allison M, Emerson A, Ramaswamy M. Retention strategies in working with justice-involved women. *J Correct Health Care* 2019 Jul 01;25(3):231-237 [FREE Full text] [doi: [10.1177/1078345819853310](https://doi.org/10.1177/1078345819853310)] [Medline: [31242801](https://pubmed.ncbi.nlm.nih.gov/31242801/)]
44. Hightow-Weidman L, Muessig K, Knudtson K, Srivatsa M, Lawrence E, LeGrand S, et al. A gamified smartphone app to support engagement in care and medication adherence for HIV-positive young men who have sex with men (AllyQuest):

- development and pilot study. *JMIR Public Health Surveill* 2018 Apr 30;4(2):e34 [[FREE Full text](#)] [doi: [10.2196/publichealth.8923](https://doi.org/10.2196/publichealth.8923)] [Medline: [29712626](https://pubmed.ncbi.nlm.nih.gov/29712626/)]
45. Shah P, Madhiwala N, Shah S, Desai G, Dave K, Dholakia N, et al. High uptake of an innovative mobile phone application among community health workers in rural India: an implementation study. *Natl Med J India* 2019;32(5):262. [doi: [10.4103/0970-258x.295956](https://doi.org/10.4103/0970-258x.295956)]
46. Lau Y, Htun TP, Tam WS, Klainin-Yobas P. Efficacy of e-technologies in improving breastfeeding outcomes among perinatal women: a meta-analysis. *Matern Child Nutr* 2016 Jul;12(3):381-401 [[FREE Full text](#)] [doi: [10.1111/mcn.12202](https://doi.org/10.1111/mcn.12202)] [Medline: [26194599](https://pubmed.ncbi.nlm.nih.gov/26194599/)]

## Abbreviations

**ANM:** auxiliary nurse midwife  
**ASHA:** Accredited Social Health Activist  
**BEST4Baby:** Breastfeeding Education Support Tool for Baby  
**CHW:** community health worker  
**EBF:** exclusive breastfeeding  
**IMR:** infant mortality rate  
**LMIC:** low- and middle-income country  
**mHealth:** mobile health  
**PC:** peer counselor  
**SCT:** Social Cognitive Theory  
**SUS:** System Usability Scale  
**TJU:** Thomas Jefferson University  
**TPB:** Theory of Planned Behavior  
**UCD:** user-centered design

*Edited by A Mavragani; submitted 11.08.21; peer-reviewed by M Herron, B Chaudhry, B Nieves Soriano, F Yu; comments to author 13.11.21; revised version received 05.03.22; accepted 26.05.22; published 08.09.22.*

### *Please cite as:*

Ma T, Chang K, Alyusuf A, Bajracharya E, Washio Y, Kelly PJ, Bellad RM, Mahantashetti NS, Charantimath U, Short VL, Lalakia P, Jaeger F, Goudar S, Derman R

*Design, Development, and Testing of BEST4Baby, an mHealth Technology to Support Exclusive Breastfeeding in India: Pilot Study* *JMIR Form Res* 2022;6(9):e32795

URL: <https://formative.jmir.org/2022/9/e32795>

doi: [10.2196/32795](https://doi.org/10.2196/32795)

PMID: [36074546](https://pubmed.ncbi.nlm.nih.gov/36074546/)

©Tony Ma, Katie Chang, Amal Alyusuf, Elina Bajracharya, Yukiko Washio, Patricia J Kelly, Roopa M Bellad, Niranjana S Mahantashetti, Umesh Charantimath, Vanessa L Short, Parth Lalakia, Frances Jaeger, Shivaprasad Goudar, Richard Derman. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 08.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# An Intervention Offering Self-management Support Through mHealth and Health Coaching to Patients With Prostate Cancer: Interpretive Description of Patients' Experiences and Perspectives

Louise Faurholt Obro<sup>1,2</sup>, RN, MSc; Palle Jörn Sloth Osther<sup>1,2</sup>, MD, PhD; Jette Ammentorp<sup>3,4</sup>, RN, MSc, PhD; Gitte Thybo Pihl<sup>5</sup>, RN, MSc, PhD; Kasper Kvolts Heiselberg<sup>1,6,7</sup>, MSc; Peter Gall Krogh<sup>6</sup>, MSc; Charlotte Handberg<sup>8,9</sup>, RN, MPH, PhD

<sup>1</sup>Urological Research Center, Vejle Hospital (Lillebaelt Hospital), University Hospital of Southern Denmark, Vejle, Denmark

<sup>2</sup>Department of Regional Health Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark

<sup>3</sup>Centre for Research in Patient Communication, Odense University Hospital, Odense, Denmark

<sup>4</sup>Department of Clinical Research, University of Southern Denmark, Odense, Denmark

<sup>5</sup>UCL, Vejle, Denmark

<sup>6</sup>Department of Engineering, Aarhus University, Aarhus, Denmark

<sup>7</sup>Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark

<sup>8</sup>National Rehabilitation Center for Neuromuscular Diseases, Aarhus, Denmark

<sup>9</sup>Department of Public Health, Faculty of Health, Aarhus University, Aarhus, Denmark

**Corresponding Author:**

Louise Faurholt Obro, RN, MSc

Urological Research Center

Vejle Hospital (Lillebaelt Hospital)

University Hospital of Southern Denmark

Beriderbakken 4

Vejle, 7100

Denmark

Phone: 45 24994502

Email: [louise.faurholt.obro@rsyd.dk](mailto:louise.faurholt.obro@rsyd.dk)

## Abstract

**Background:** Observational management strategies such as active surveillance and watchful waiting are considered to be acceptable approaches in patients with low-risk localized prostate cancer and a safe alternative to aggressive treatment. During observational management, treatment is postponed until the disease progresses, which often never occurs. However, approximately 90% of patients with a low-risk disease choose aggressive treatment owing to anxiety. Strategies to address anxiety are needed for optimal management of this population and to improve the quality of life of patients with low-risk localized prostate cancer. A review highlighted that mobile health (mHealth) in tandem with health coaching can support patients' self-management of health behaviors and improve well-being.

**Objective:** This study aims to explore patients' experiences with and perspectives on an intervention offering self-management support through the use of mHealth devices and health coaching to identify supportive features that enable patients to perform sustainable changes that improve well-being.

**Methods:** We used an interpretive description approach, combining semistructured interviews with 13 purposively selected patients with prostate cancer and participant observations of patient-coach interactions in coaching sessions. The interviews were transcribed and analyzed. The self-determination theory was used as a theoretical lens. Field notes and coaching notes from each session were used to orient data generation and confirm or challenge the analysis.

**Results:** Our analysis suggested that patients' self-awareness and psychological identity influenced their experiences with and perspectives on the self-management support offered by mHealth and health coaching in clinical practice. The patients' individual experiences and perspectives indicated that they placed themselves in a dynamic continuum of sustaining or repressing their identity, self-awareness, and individual qualities. Our analysis revealed 4 interacting themes, all related to the psychological identity of the patients.

**Conclusions:** For the group of patients with prostate cancer to experience well-being, we found it important for them to sustain their self-image when offered a self-management intervention. Motivation and autonomy were important aspects for the individual patients to sustain their self-image throughout the intervention. In contrast, demotivation and a sense of paternalism could result in fostering an experience of having to repress self-awareness.

(*JMIR Form Res* 2022;6(9):e34471) doi:[10.2196/34471](https://doi.org/10.2196/34471)

## KEYWORDS

mobile phone; mobile health; mHealth; prostate cancer; self-management; health coaching; coaching

## Introduction

### Background

Prostate cancer (PCa) is the most common malignancy among men [1]. Most men with PCa have low-risk localized PCa, and the cancer is unlikely to become life threatening [2]. Aggressive treatment may provide little survival benefit for patients with low-risk localized PCa while increasing the risk of side effects, such as erectile dysfunction, urinary incontinence, fatigue, weight gain, and depression that may severely impact the quality of life [2]. Observational management strategies such as active surveillance (AS), which is used in a curative setting, and watchful waiting, which is used in a palliative setting, have been shown to be feasible in patients with low-risk localized PCa [3]. Low-risk localized PCa is seen as a chronic disease in which treatment is postponed until the disease progresses, which often never occurs [3]. The observational management strategy involves lifelong plasma prostate-specific antigen testing and, in some cases, repeated biopsies of the prostate [3]. Another important aspect of observational management is to encourage patients to live a healthy and active life because alcohol, smoking, and obesity are among the risk factors for cancer progression [4]. Although observational management appears to be a safe alternative to aggressive treatment, approximately 90% of patients with a low-risk disease choose aggressive treatment [5]. The Prostate Cancer Research International Active Surveillance (PRIAS) study by Bokhorst et al [5] revealed that approximately 13% of patients on AS choose to initiate active treatment for PCa owing to anxiety. Thus, as argued by Bokhorst et al [5], strategies to address anxiety are needed for optimal management and to improve health behaviors and quality of life in patients with low-risk localized PCa. Well-being includes having good mental health, high life satisfaction, a sense of meaning or purpose, and the ability to manage stress and anxiety [6]. Self-management strategies represent a promising approach for treating chronic conditions and improving well-being and quality of life [6].

Mobile health (mHealth) devices, such as smartphones, fitness trackers, and wearables, represent a new generation of tools with the potential to improve patient self-management [6]. mHealth provides reliable and safe data collection outside the clinical setting and facilitates the delivery of interventions (eg, instruction in behavioral change) [6]. Although the promise of mHealth seems appealing, some challenges were highlighted by Woods et al [7], who found that patients often have a significant barrier to using mHealth in everyday life. They found that health care professionals are requested to support patients in adopting technological devices to ease the integration of

mHealth as part of self-management strategies [7]. A recent scoping review indicated that mHealth and health coaching work synergistically and enhance patients' self-management [8].

### Objectives

Health coaching is a patient-centered intervention in which a health coach guides a patient in making behavioral changes. This is typically achieved by encouraging patients' active participation in self-management based on personal objectives and individual motivational readiness to change [9]. Palmer et al [10] defined health coaching as "a practice of health education and health promotion within a coaching context, to enhance the well-being of individuals, and to facilitate the achievement of their health-related goals." Accordingly, combining mHealth and health coaching may be a promising approach to support self-management in patients with PCa in observational management. Enabling and empowering patients to assume a more active role in the management of their disease has been shown to increase their quality of life and lower distress [11,12]. Growing evidence suggests that self-management strategies can benefit patients [11,13]. Self-management can be defined as the initiatives undertaken by individuals to promote their health and well-being [14]. It includes the actions individuals take toward a healthy lifestyle: managing their lifelong disease, management of their emotional health and well-being, and prevention of further illness [11,14]. Although several studies point to self-management as a promising strategy for patients to manage their chronic cancer disease [11,14], others address a key issue that needs to be discussed: how health care professionals can support self-management in an evidence-based, structured way and how self-management support can be integrated into clinical practice [15]. Thus, the objective of this study was to explore patients' experiences with and perspectives on an intervention offering self-management support through different mHealth devices and health coaching to investigate what supported them in making sustainable changes that improved their well-being.

## Methods

### Setting and Sampling

All participants were recruited from the urological outpatient clinic at Vejle Hospital, a part of the Lillebaelt Hospital, Vejle, Denmark. Participants were informed and invited to participate in the study at the urological outpatient clinic by their urologist. We included patients with PCa on AS or watchful waiting who could read and speak Danish. A total of 13 participants were recruited and divided into 2 groups: group 1 between June and

August 2017 and group 2 between April and August 2020 (Table 1). The participants were included using purposive sampling [16]. In addition, 4 female urological nurses were included in the study to coach the participants (Table 2). The nurses had completed a 2-day course in health coaching given by a certified health coach.

**Table 1.** Characteristics of the 13 participants including their mobile Health (mHealth) devices and tracked activity.

Group number	Participant number	Recruitment period	Age (years)	Ethnicity	Employment status	Time since diagnosis (years)	mHealth devices	Tracked activity
1	1	June 2017 to August 2017	71	Danish	Retired	5	BTTN <sup>a</sup>	Water intake
1	2	June 2017 to August 2017	70	Danish	Retired	5	BTTN	Water intake
1	3	June 2017 to August 2017	72	Danish	Retired	3	BTTN	Water intake
1	4	June 2017 to August 2017	68	Danish	Self-employed	1	BTTN	Water intake
1	5	June 2017 to August 2017	82	Danish	Retired	6	BTTN	Water intake
1	6	June 2017 to August 2017	70	Danish	Retired	3	BTTN	Water intake
1	7	June 2017 to August 2017	69	Danish	Employed	2	BTTN	Water intake
2	8	April 2020 to August 2020	80	Danish	Retired	2	Fitness tracker and music device	Steps and well-being
2	9	April 2020 to August 2020	81	Danish	Retired	5	Fitness tracker and music device	Steps and well-being
2	10	April 2020 to August 2020	75	Danish	Retired	2	Fitness tracker and music device	Steps and well-being
2	11	April 2020 to August 2020	78	Danish	Retired	3	Fitness tracker and music device	Steps and well-being
2	12	April 2020 to August 2020	72	Danish	Retired	3	Fitness tracker and music device	Steps and well-being
2	13	April 2020 to August 2020	75	Danish	Retired	3	Fitness tracker and music device	Steps and well-being

<sup>a</sup>BTTN: Bluetooth button with connection to the *My Course* app and the electronic patient journal.

**Table 2.** Characteristics of the 4 nurse coaches.

Group number	Nurse number	Recruitment period
1	1	June 2017 to August 2017
1	2	June 2017 to August 2017
2	3	April 2020 to August 2020
2	4	April 2020 to August 2020

## Intervention

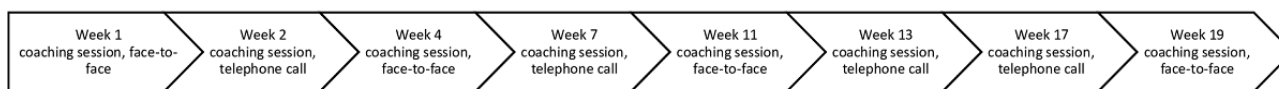
All 13 patients participated in a 19-week program that included 8 individual coaching sessions: 4 face-to-face visits to the outpatient clinic, lasting between 45 and 60 minutes each, and 4 telephone calls of 30 minutes each (Figure 1). Participants met the same coach throughout the program. The program aimed

to provide ongoing support and guidance for participants to set goals and sustainable objectives and make changes that improved overall health and well-being. In the first coaching session, the participants received an mHealth device that tracked the activity of interest. The tracked activities were related to overall health, well-being, or both. Those related to overall health were physical activity or water intake; they were chosen

because they are simple to track and are important aspects of living a healthier life [6]. The tracked activities related to well-being were self-reflection on one’s life, which was chosen

because self-reflection is important for good mental health and dealing with emotional challenges [6].

**Figure 1.** The 19-week coaching program.



After the first session, all coaching sessions started with the nurse coach and a participant jointly evaluating the data. In addition, difficulties with the mHealth devices were addressed at the beginning of each session. After evaluating the data, the nurse and the participant talked about facilitators and barriers related to the participants’ individual goals and optionally setting new goals. At the end of the session, the nurse summarized the session.

**Group 1**

All participants in group 1 felt that water intake could be improved and agreed to register daily water intake with a Bluetooth button (Figure 2) connected with the My Course app and the electronic patient journal. The participant tracked by a push on the Bluetooth button every time they drank a glass of water. During the following sessions, the coach and the participant evaluated the assessment and discussed the barriers and motivation for achieving the goal of increased water intake.

**Figure 2.** Picture of the mobile health devices.



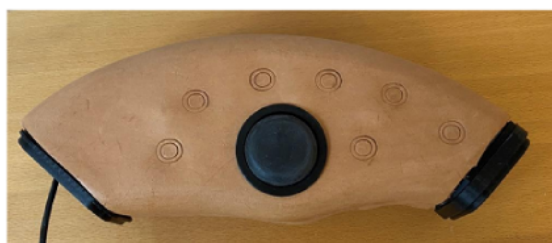
Bluetooth button bt.tn



FitBit Charge 4 activity tracker



Garmin Vivosport activity tracker



Music device

**Group 2**

In the first session, in group 2, the coach and the participants used the *Wheel of Life* [17] to identify and set objectives. The *Wheel of Life* by Meyer [17] is a visual model used in coaching to help clients understand their current sense of balance or

fulfillment in life [17]. The wheel, divided into pies, usually consists of 8 to 10 client-identified categories considered important for a full or balanced life [17]. Participants rated their level of satisfaction within each category and then mapped them onto an image of a wheel. This provided them with an immediate

overview of their current *life balance* [17]. Subsequently, each category was rated from 1 (not at all satisfied) to 10 (completely satisfied or could not be improved) by drawing a line across the corresponding number in that section of the pie and then shading below it. Completing this leaves the participant and the coach with a jagged wheel that illustrates the areas for growth. Following this exercise, the participants in group 2 were equipped with 2 mHealth devices: a fitness tracker (FitBit or Garmin; Figure 2) and a music device (Figure 2). The participant tracked their steps using a fitness tracker. The objective of the use of a fitness tracker was for the participants to reflect on their daily activities. The objective of the music device was to improve the participants' well-being by spurring self-reflection. The music device was intended to function as a digital diary in which participants were asked to play a short musical phrase, a fragment of a tune, or something similar that reflected their well-being and emotions on a daily basis. The coach and participants listened to the recordings during the coaching sessions and then discussed what was recorded.

### Methodological Approach

On the basis of the need to explore patients' experiences with and perspectives on an intervention offering self-management support through mHealth and health coaching and the need to inform and create knowledge in collaboration with clinical practice, we applied the qualitative inductive methodology of interpretive description (ID) [16]. ID was chosen as the methodology because it addresses complex explorative clinical questions while producing practical outcomes [16]. ID applies the notion that social influences are formed by people and from people and their actions; in contrast, it also seeks a nuanced understanding of an individual's perceptions of the phenomenon of interest [16]. The methodology draws on other qualitative studies such as ethnography, grounded theory, and phenomenology. Furthermore, ID stresses the value of a *research logic*, permitting the researcher to apply and combine the methods needed to fully answer the research question [16]. The flexibility of ID is practical when exploring a field in which unexpected findings may occur, which would require an adjusted strategy [16]. ID allows for early (preliminary) analysis of data that enables mutual adjustment of the data being collected and the models applied for analysis [16]. ID invites a theoretical framework to help structure the study of data material at hand, for example, interview guide, analysis, and discussion [16], and therefore, we also used the self-determination theory (SDT) [18]. We used SDT because it points to how the achievement of health-related goals and well-being is more effective and lasting when patients are autonomously motivated [18]. Patients' motivation depends on their personal convictions and the degree to which their psychological needs for competence, autonomy, and relatedness are fulfilled [18]. Furthermore, we used SDT to construct the interview guide and conduct interviews and analysis to capture the impact of the mHealth device and health coaching on patients' psychological needs and motivation.

### Data Collection

Data consisted of semistructured interviews in which a few predefined questions with an open approach were asked, for example, "How did you experience the use of mHealth?" and

"What are your perspectives on coaching?" Data also consisted of field notes by the first author (LFO) reporting on observational studies [16] of the coaching sessions and the coaches' notes from each session, totaling approximately 90 pages. After each coaching session, the coaches wrote notes concerning the patient's progress, patient's experiences with the intervention, and possible tasks the patient had opted for. A total of 26 semistructured interviews were conducted by the first author, 13 before the intervention and 13 at the end of the study. The participants were interviewed in the clinic or by phone, and the interview lasted between 30 and 70 minutes. The interviews were audiotaped and transcribed verbatim.

Data from the observations were used to provide insight into coach-participant interactions and to support and elaborate interview data. The observations were made in October 2017 and October 2020 (approximately 5 hours of observations).

### Data Analysis

The analysis was guided by the ID methodology [16]. Interview data were transcribed, anonymized, and transferred to the QRS NVivo software, version 12. In the first step of the analysis, the first author (LFO) and the last author (CH) read the data and developed an initial coding structure based on the initial analysis. LFO and CH then performed a more specific coding strategy, shifting between the process of coding and taking *astep back* to gain perspective on the data material as a whole. After the initial coding, LFO and CH further refined, described, and discussed themes grounded in the remaining data. This step was repeated, ensuring that the themes comprised the data, and subsequently, the themes were discussed with the rest of the research team (KH, PGK, JA, and GTP). Correspondingly, we addressed any inconsistencies within and between the interview data, field notes, and notes from the coaches. Finally, we created a model that represented the analytical findings in a hierarchy [16].

### Ethical Considerations

All participants received oral and written information on the purpose and methods of the study, including their right to withdraw at any time. If participants agreed to participate, they signed an informed consent form. All data were anonymized and stored in a secure place that was approved by the Danish Data Protection Agency in accordance with the General Data Protection Regulation. This study was approved by the Regional Committees on Health Research Ethics for Southern Denmark (case ID: 20212000-105).

## Results

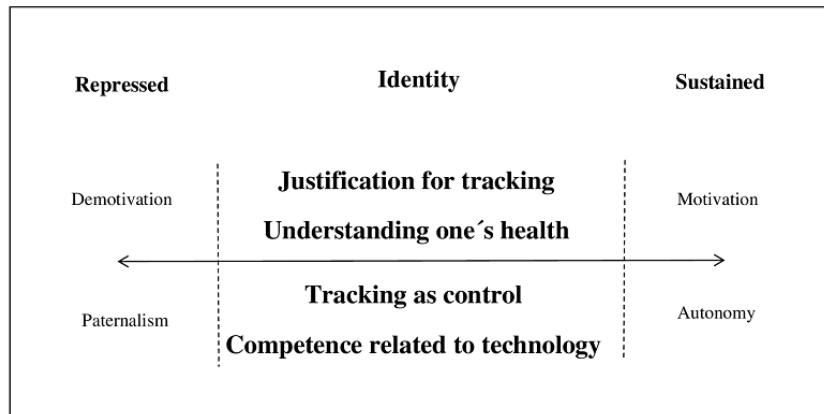
### Overview

Our analysis revealed that patients' identity, including their self-awareness and self-image, seemed to influence their experience with and perspective on self-management support through mHealth and health coaching in clinical practice. The patients' individual experiences and perspectives indicated that they placed themselves in a dynamic continuum of sustaining or repressing their sense of identity. Our analysis revealed four interacting themes related to the identity and self-awareness of the patients (Figure 3): (1) justification for tracking and (2)

understanding one's health collectively shaped a spectrum of patient views ranging from being motivated for self-tracking to being demotivated and (3) tracking as control and (4)

competence related to IT ranged in a spectrum from a sense of autonomy to paternalism. An illustration of the findings is presented in Figure 3.

**Figure 3.** Understanding the perspectives of patients with prostate cancer on an intervention offering self-management support through mobile health and health coaching in clinical practice.



### Justification for Tracking: Motivation and Demotivation

The patients explained their tracking of the intervention by unfolding a spectrum from being motivated to being demotivated. Some patients stated that they were motivated to track the data because they liked the idea of self-tracking. However, some patients explained that they did not find tracking motivating but still tracked data because they liked the interaction with the coach:

*I needed to talk to someone about it (the stressful situation at home) and the coach was really nice to talk to, in a sense..., she could see things differently. [P13]*

The patients who found tracking motivating related it to an opportunity to gain insight into the intervention. One patient explained the following:

*The reason I signed up for this was that I got a request, and then I'm probably the kind of person who..., if you can do something to gain insight into how you can manage things better in the future, health care and medical related things, then I'm of the opinion that it is my civic duty to sign up. [P10]*

During the coaching sessions, some participants enthusiastically demonstrated their devices to the coach and described how some technological aspects could be improved. The coach expressed appreciation for the patients' contributions, which seemed to motivate them even further.

Some patients described that they were tracking because of a sense of duty, either toward the health professionals or their fellow patients with PCa. The sense of duty was explained by some patients as a motivation, whereas for others, it felt demotivating.

A feeling of demotivation for tracking emerged among patients who were struggling to understand the aim and logic behind the intervention. When they could not comprehend the logic behind

the devices, in particular the music device, they expressed feeling demotivated about using the device or only using it to repress a feeling of guilt.

One participant using the music device explained the following:

*It gave me nothing. On the contrary, it became a pestilence for me to track every day, because I did not benefit from it. I was just trying to get some little tunes and stuff like that; well, nothing...but the fact that it could provide me with a little feeling of calmness and relaxing, right? [P11]*

From the quotes and the observations, it appeared that when the participants experienced themselves and their insights as important, they experienced that their self-image was sustained, and they felt motivated.

In contrast, other patients argued that when they could not contribute with valuable insights and felt that they were merely "a bystander" (P9) in the intervention, and thereby were only tracking data because of feelings of guilt and duty, they experienced repression of their self-image and were demotivated.

### Understanding One's Health: Motivation and Demotivation

Several patients described that participating in the intervention provided them with the ability to better understand their health, which they found motivating. Most of the participants in the group that tracked water intake explained that a balanced water intake was related to their well-being and was a motivational factor for them. The patients in this group expressed that by gaining insight into health through tracking and coaching, they adopted a better rhythm in life. They explained that they became more aware of drinking water throughout the day, instead of drinking the whole amount at the end of the day, entailing frequent toilet visits during the night. One patient explained the following:

*I think that I benefitted a lot from participating in the project. It has been great, because I have obtained a*



*stable rhythm, remembering to drink water, which is important when you are my age, right? So, yes it has. It has given me a lot, I would say, not just a little.* [P2]

It seemed that when the patients had a successful understanding of the importance of water intake, they were often more motivated to set a new health objective for increasing balance and well-being in their lives. A patient explained the following:

*After attending this intervention my wife and I became more aware of drinking more water and eating healthier, and this gave us a lot of energy to go for a walk or two.* [P7]

Some patients also highlighted the importance of interacting with the coach to gain a better understanding of their health. The patient described that the coach helped them translate their tracking data into “a language that I understand” (P6). The coach seemed to play an important role for the patients, turning insights from the data into advice for the patients and keeping them motivated.

The importance of the coach in fostering understanding and motivating the patients can also be seen in the observations. In a coaching session, a patient (P9) seemed depressed and resigned when the coach asked questions about their health and life. During the session, the patient experienced anxiety because of the PCa treatment. He told the coach that he did not trust the urologist to control his illness. The coach wrote the following in her notes: “He (P9) seems nervous for his prostate cancer. Also, occasionally explains thoughts about what it would be like not to live anymore” (C2). The coach also described in her notes that, at every session, she engaged in conversation with him on his history of PCa and treatment. She said that he seemed to be increasingly positive and motivated after each session. During the last interview, the patient explained the following: “The conversations I had with (C2), have given me a positive view on life” (P9). As illustrated in these quotes and observations, some patients were motivated by the conversational support from the coach.

Some patients explained that through the understanding of their health and increased self-awareness gained by participating in the intervention, they were able to sustain their self-image and had more courage to ask questions in the consultations with the urologist in the outpatient clinic. For instance, some patients explained that they were more motivated to ask their urologist questions because they had gained a better understanding of their health. Furthermore, the patients explained that registrations on the mHealth devices could contribute to discussions about nocturia with their urologists. In addition, the patients explained that the tracking provided insights for the urologist into the patients’ everyday lives, which could help individualize their treatment.

However, some patients described the newfound understanding of their health as demotivating. The tracking of their data could, for some, be an unpleasant reminder and awareness of their illness and being ill. Furthermore, it appeared that some participants neglected their disease. In this context, a coach described in her notes that a few patients did not believe that they had been diagnosed with PCa, causing the coach to look

up the patients’ journals to confirm the PCa diagnosis. Some patients explained that talking about their illness and health had a negative psychological effect on them. Regarding this, some patients expressed that they did not think that self-tracking and coaching could contribute to any new understanding; conversely, they feared that they would be confused by the tracking data and by talking to the coach.

### Tracking as Control: Autonomy and Paternalism

Some patients reported that they had experienced tracking as a method of control. They described a feeling of control, ranging from autonomy to paternalism. Having control over one’s life and maintaining autonomy are important. The feelings of control and autonomy appeared to be linked to each patient’s previous working life and self-image. Some explained that they held high positions and served important functions in their working lives with high degrees of decision integrity. For these patients, PCa diagnosis led to a loss of control. One patient explained the following:

*I’m old policeman and I’ve had have had many employees, and stuff like that. Taken a lot of beatings and so on. This prostate cancer diagnosis fucked me up a little. That was actually the worst part. It got personal.* [P3]

The loss of control and autonomy for some patients seemed to impact their well-being negatively. However, some patients described self-tracking as a way to regain autonomy and well-being. A patient explained the following:

*I really thought I was getting enough exercise during the day, but after I have begun wearing the watch (activity-tracker), I could suddenly see that I did not. Now I can see how many steps I have taken, and then I can judge for myself whether I should go for a long or short walk.* [P11]

The importance of control and autonomy was also observed during the coaching sessions. In one session, a patient (P8) stated that he had been a lawyer and had always been in control. He emphasized his ability to make decisions for himself regarding data tracking. He appeared relaxed when talking with the coach and explained that he liked the idea of self-tracking and that he saw it as a way for him to control his body.

With the use of mHealth, some patients described that they had gained the ability to make judgments to do something good for themselves. As a result of the intervention, the patients seemed to feel more in control of their behavior and how this control could impact their lives. Furthermore, it appeared that when the use of self-tracking was experienced as meaningful and the patients were tracking out of their own free will without external pressure, they explained a sense of autonomous control, which seemed to increase their well-being.

For some patients, self-tracking could be regarded as an external control, which could lead to a sense of being subjected to paternalism. These patients explained that they experienced self-tracking as a way of being controlled by the coach. In a coaching session, a patient (P12) explained that he could not see how he could benefit from tracking data and that he did not want to wear the fitness tracker. He told the coach that he had

his own device at home and that he would use it instead. Likewise, other patients stated that they saw self-tracking as a way for the coach to push them to do something they did not want to do. However, some patients explained that at the beginning of the intervention, they had a sense of being forced to track data but experienced that the tracking became more autonomous at the end of the intervention. A patient said the following:

*I did it to satisfy the coach. She told me to try recording small melodies. I felt like I had to do it...but in the end it might have given me something. [P10]*

It appeared that self-tracking could bring autonomy, control of one's life, and increase well-being when tracking was experienced as meaningful and without an external paternalistic control.

### Competence Related to Technology: Autonomy and Paternalism

The patients described how competencies related to technology ranged from experiencing autonomy to paternalism. All patients had their own smartphone and other electronic devices such as tablets, computers, or smartwatches. Many patients said that they used their smartphone or other devices for several hours each day. The patients explained that they used their smartphone to text, help with wayfinding, read the newspaper, and watch movies. When patients rated their competence level in relation to technology, they ranged from limited skills to advanced user. The patients in the group that tracked activity with the fitness tracker explained that the device was easy to use. A participant mentioned the following:

*You do not need any special technological skills to use it, it can be used by anyone, you only need to put it on your hand and off you go. [P7]*

The patient (P7) also described that he found the fitness tracker suitable for him because he felt that he had limited competence in mastering IT. He explained that it was important for him to have the right match between the required level of competencies and skills needed to operate the device and that the right match was the reason that he kept tracking.

In addition, some patients described an experience of autonomy related to having *the right* IT competencies. Some patients described that they had worked with technology and had thereby achieved IT competencies. The patients explained that when they experienced having the right level of IT competence, they found the device easier to use and were more likely to experiment with the devices. A patient said the following:

*I am used to working with technology and find it exciting. So, I tried to place the cable differently and then it worked. [P6]*

It seemed that the experience of sustaining one's self-image and autonomy occurred if the patients had prior and adequate competencies to use the mHealth devices and found the mHealth devices easy to use.

However, some patients did not feel competent to use these devices. During a coaching session, a patient (P3) expressed disinterest in using the device. He told the coach that he could

not get the Bluetooth button to work and that he could not read the instructions because they were in English, so he returned the button to the coach. Likewise, a patient, in an interview, said that he became frustrated and insecure when he had to change the battery in the Bluetooth button, fearing that "it would explode" (P1). The sense of not having the proper competencies to use the mHealth devices seemed to frustrate the patients. It appeared that a feeling of being subject to paternalism could occur if the patients were forced to use the devices and did not possess the proper competencies to use them.

Some of the patients explained that they only used the devices to please the coach, and other patients found the mHealth device too abstract, particularly the music device. These patients explained that they would have liked to obtain more information about how to use the music device and the purpose of using it. In addition, a patient explained that he felt powerless because he could not see the point of using either the fitness tracker or the music device:

*Maybe the fitness tracker supports being able to keep up with how much you walk, but I cannot see the point of having to play the music device in the evening. So, I did not use it. [P4]*

It seemed that patients experienced autonomy when they experienced possessing the required IT competencies. Consequently, when patients were unable to operate the devices meaningfully, they experienced reduced autonomy.

## Discussion

### Principal Findings

The objective of this study was to explore the experiences of patients with PCa and their perspectives on an intervention offering self-management support through mHealth and health coaching aimed at helping the patients achieve well-being. We applied the SDT by Ryan and Deci [18] as a theoretical framework to interpret and discuss our findings. Our analysis indicated that patients' identity and self-image influenced their experiences with and perspectives on the self-management afforded by mHealth and health coaching in clinical practice. When the patients experienced having the suitable skills for the intervention, they were able to sustain their self-image, and then, they experienced well-being. This finding is in accordance with the SDT by Ryan and Deci [18,19]. Ryan and Deci [18] argued that the sense of self plays a significant role when new tasks and actions are required in new social contexts, and when self-awareness and esteem match the tasks at hand, people feel satisfied and well. Our findings are consistent with the identity-based motivation theory, which highlights that humans are motivated if they experience that their self-image fits with their current situation and circumstances and that humans prefer to act in ways that lead to sustaining their self-image rather than repressing it [20]. When the patients in this study experienced that the tasks and actions in the intervention fitted their identity, they felt that the tasks and actions became more meaningful to them, and their identity was sustained [18,19]. Moreover, when the tasks and actions did not align with the task at hand, the patients tried to steer toward not having to repress their identities. This caused some patients to reject using the devices

or use their private devices [18]. According to Ryan and Deci [18], the fulfillment of autonomy, competence, and relatedness is essential for patients to achieve well-being. Our study indicates that each patient's sense of self could be sustained if their basic psychological needs were fulfilled, which seemed to motivate the patients to participate in the intervention [18]. Our findings highlight that the patients seemed to be more motivated and satisfied when their actions were self-initiated; however, we also found that the patients were demotivated if they experienced being directed (rather than motivated) by the coach. Franklin et al [21] supported these findings and argued that to facilitate well-being in a self-management intervention, an autonomy-supportive environment is important to increase intrinsic motivation for the sustained self-regulation of health behavior. Furthermore, research has illuminated how autonomy is essential for achieving self-management goals and for improving health-related goals [22]. Our findings also highlight that the patients' level of health literacy (eg, individual skills and technical competencies) should be taken into account when offering a self-management intervention through mHealth for the patients to successfully engage in self-tracking [18,23]. Patients face new tasks and require a new set of skills when using mHealth, which carries the additional challenge of digital health literacy [23]. As a consequence of the use of mHealth, the health care system risks excluding some patients [23]. This consequence was demonstrated in our study by some patients who declined to use the mHealth device when they could not understand the logic and rationale behind its use. In line with this, several studies have concluded that objective self-tracking data, such as activity or heart rate, are easy to capture and that patients understand and can convert the understanding of such data into actions that lead to healthy behavior [24]. This emphasizes the importance of the patients' health literacy level in their understanding of the logic and rationale behind the tracking and, in accordance with this, making healthier choices.

Hilty et al [25] claimed that not all patients may be suitable for mHealth, and research shows that patients with lower health literacy are less likely to use different types of digital health tools than those with high health literacy. To include more patients in self-management interventions through mHealth and health coaching, the health care system must consider the patients' different needs and skills when offering mHealth devices as a part of their care. Therefore, to facilitate a successful self-management intervention, it could be important to screen patients' IT competencies before offering them an mHealth device, as Chan et al [26] suggested. Health care professionals are another pivotal element in the implementation of mHealth devices because they are important for the initiation and facilitation of self-management interventions during their interaction with patients [27]. In our study, we also found that the patients' need for relatedness with the nurses (coaches) had implications on their well-being, and the patients expressed that it was important to have a feeling of belonging and being significant in the eyes of the nurses [18]. In this way, the role of the coach was important and motivating for the patients to translate data into everyday health-related advice and support for behavioral change. These findings are supported by a randomized trial that confirmed the importance of the interaction between patients and coaches as crucial in improving the

well-being of patients [28]. This also correlates with a recent study showing that nurse-led coaching increases emotional well-being in patients with chronic diseases [29].

### Study Strengths and Limitations

As this study was designed within the framework of the ID methodology, as outlined by Thorne [16], it ensured that the research was both methodologically and interpretively rigorous. Methodological rigor and interpretive rigor are described as the 2 types of rigor required for qualitative research to be considered credible [30]. Epistemological integrity was achieved by ensuring that the design and implementation of this study were consistent with ID [16] and SDT. Credibility was enhanced through researcher triangulation. The author team consisted of people with various backgrounds in health care and research, which provided various perspectives on the analysis, interpretation, and understanding of the data.

Moreover, the credibility of our findings was enhanced through the triangulation of interview data and observational data, which strengthened analytical interpretations owing to the variety of insights [30]. We used participant observations and notes from the coaches to provide insights and challenge the interviews with the patients and to help interpret the patients' experiences with and perspectives on the self-management intervention, which could have prevented placing an overemphasis on the individual interviews [31]. For example, the notes from the coaches provided unique insights regarding the interaction between the patients and the coaches and, in this way, added to the information gained through the interviews. The notes provided insight into the fact that the patients seemed more honest with the coach than with the interviewer. This may have meant that the opinions expressed by the patients during the interviews were more positive, but combined with the observational data, the interview data provided us with realistic insights into the patients' perceptions of using mHealth. Overall credibility was sought through the use of consistent analytic logic [16]. The use of the same interview guide for all 26 interviews allowed the same structural and analytical consistency.

A limitation of the study is that it is unclear whether we examined the use of mHealth, health coaching, or both. Furthermore, a limitation could be that the health coaching was conducted in a room in the outpatient clinic, and the physical structures and social dynamics could have affected the interaction between the coach and the participant [32]. Research has pointed out that the environment in which coaching takes place is significant for facilitating a coaching conversation [32], and as our coaching was performed in the outpatient clinic, this could have resulted in a more traditional medical conversation, meaning that the potential of the coaching conversation was not fully used. Although purposive sampling does not confer transferability, it has provided in-depth information about patients with PCa in observational management and their experiences with and perspectives on the intervention [16]. On the basis of our description of the research context and our assumptions, we believe that our findings can be applied to patients with other cancer diseases in other health care settings in Western countries.

## Conclusions

For patients with PCa to achieve well-being, we found it important for them to be able to sustain their identity and self-awareness when offered a self-management mHealth intervention. The patients' individual experiences and perspectives indicated that they placed themselves in a dynamic continuum of sustaining or repressing their identities and self-images. For the patients to sustain their self-image throughout the self-management intervention, motivation, autonomy, having the suitable competencies, and the interaction with the nurse coach appeared to be important aspects. In

contrast, demotivation and paternalism could result in an experience of having to repress their identity and self-image, and thus, they have a negative effect on the use of mHealth. The patients seemed to be motivated to track when they felt in control, had the suitable competencies and skills related to technology, and when they experienced that their insights were significant to others.

To further understand the barriers and potential for developing a successful self-management intervention, future research should investigate health care professionals' experiences with and perspectives on the use of mHealth and health coaching.

## Acknowledgments

The authors would like to thank all participating patients and the staff of the urological outpatient clinic at Vejle Hospital.

## Authors' Contributions

All authors have made substantial contributions to the following: (1) the conception and design of the study, acquisition of data, or analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; and (3) final approval of the version to be submitted.

## Conflicts of Interest

None declared.

## References

1. Siegel R, Miller K, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019 Jan;69(1):7-34 [FREE Full text] [doi: [10.3322/caac.21551](https://doi.org/10.3322/caac.21551)] [Medline: [30620402](https://pubmed.ncbi.nlm.nih.gov/30620402/)]
2. Aminsharifi A, Simon R, Polascik TJ, Robertson CN, Sudan DL, Collins BH, et al. Evaluation and active treatment versus active surveillance of localized prostate cancer in renal transplant patients in the era of low and very low risk prostate cancer. *J Urol* 2019 Sep;202(3):469-474. [doi: [10.1097/ju.000000000000207](https://doi.org/10.1097/ju.000000000000207)]
3. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur Urol* 2017 Apr;71(4):618-629. [doi: [10.1016/j.eururo.2016.08.003](https://doi.org/10.1016/j.eururo.2016.08.003)] [Medline: [27568654](https://pubmed.ncbi.nlm.nih.gov/27568654/)]
4. Schenk JM, Neuhauser ML, Beatty SJ, VanDoren M, Lin DW, Porter M, et al. Randomized trial evaluating the role of weight loss in overweight and obese men with early stage prostate cancer on active surveillance: rationale and design of the prostate cancer active lifestyle study (PALS). *Contemp Clin Trials* 2019 Jun;81:34-39 [FREE Full text] [doi: [10.1016/j.cct.2019.04.004](https://doi.org/10.1016/j.cct.2019.04.004)] [Medline: [31002955](https://pubmed.ncbi.nlm.nih.gov/31002955/)]
5. Bokhorst LP, Valdagni R, Rannikko A, Kakehi Y, Pickles T, Bangma CH, PRIAS study group. A decade of active surveillance in the PRIAS study: an update and evaluation of the criteria used to recommend a switch to active treatment. *Eur Urol* 2016 Dec;70(6):954-960. [doi: [10.1016/j.eururo.2016.06.007](https://doi.org/10.1016/j.eururo.2016.06.007)] [Medline: [27329565](https://pubmed.ncbi.nlm.nih.gov/27329565/)]
6. Coughlin SS, Caplan LS, Stone R. Use of consumer wearable devices to promote physical activity among breast, prostate, and colorectal cancer survivors: a review of health intervention studies. *J Cancer Surviv* 2020 Jun;14(3):386-392. [doi: [10.1007/s11764-020-00855-1](https://doi.org/10.1007/s11764-020-00855-1)] [Medline: [31933148](https://pubmed.ncbi.nlm.nih.gov/31933148/)]
7. Woods LS, Duff J, Roehrer E, Walker K, Cummings E. Patients' experiences of using a consumer mHealth app for self-management of heart failure: mixed-methods study. *JMIR Hum Factors* 2019 May 02;6(2):e13009 [FREE Full text] [doi: [10.2196/13009](https://doi.org/10.2196/13009)] [Medline: [31045504](https://pubmed.ncbi.nlm.nih.gov/31045504/)]
8. Faurholt Obro L, Heiselberg K, Gall Krogh P, Handberg C, Ammentorp J, Thybo Pihl G, et al. Corrigendum to "Combining mHealth and health-coaching for improving self-management in chronic care. A scoping review" [Patient Educ Couns, 104, 4 (April 2021) 680-688]. *Patient Educ Couns* 2021 Oct;104(10):2601. [doi: [10.1016/j.pec.2021.06.035](https://doi.org/10.1016/j.pec.2021.06.035)] [Medline: [34229918](https://pubmed.ncbi.nlm.nih.gov/34229918/)]
9. Ammentorp J, Uhrenfeldt L, Angel F, Ehrensward M, Carlsen EB, Kofoed P. Can life coaching improve health outcomes?--A systematic review of intervention studies. *BMC Health Serv Res* 2013 Oct 22;13(1):428 [FREE Full text] [doi: [10.1186/1472-6963-13-428](https://doi.org/10.1186/1472-6963-13-428)] [Medline: [24148189](https://pubmed.ncbi.nlm.nih.gov/24148189/)]
10. Palmer S, Tubbs I, Whybrow A. Health coaching to facilitate the promotion of healthy behaviour and achievement of health-related goals. *Int J Health Promotion Educ* 2003 Jan;41(3):91-93. [doi: [10.1080/14635240.2003.10806231](https://doi.org/10.1080/14635240.2003.10806231)]
11. McCorkle R, Ercolano E, Lazenby M, Schulman-Green D, Schilling LS, Lorig K, et al. Self-management: enabling and empowering patients living with cancer as a chronic illness. *CA Cancer J Clin* 2011 Jan 04;61(1):50-62 [FREE Full text] [doi: [10.3322/caac.20093](https://doi.org/10.3322/caac.20093)] [Medline: [21205833](https://pubmed.ncbi.nlm.nih.gov/21205833/)]

12. Ammentorp J, Thomsen J, Kofoed P. Adolescents with poorly controlled type 1 diabetes can benefit from coaching: a case report and discussion. *J Clin Psychol Med Settings* 2013 Sep 31;20(3):343-350. [doi: [10.1007/s10880-013-9374-z](https://doi.org/10.1007/s10880-013-9374-z)] [Medline: [23900748](https://pubmed.ncbi.nlm.nih.gov/23900748/)]
13. Timmermann C, Ammentorp J, Birkelund R. Person-centred communication with cancer survivors: exploring the meaning of follow-up coaching conversations. *Scand J Caring Sci* 2021 May 24. [doi: [10.1111/scs.13007](https://doi.org/10.1111/scs.13007)] [Medline: [34028082](https://pubmed.ncbi.nlm.nih.gov/34028082/)]
14. van Riel PL, Zuidema RM, Vogel C, Rongen-van Dartel SA. Patient self-management and tracking: a European experience. *Rheum Dis Clin North Am* 2019 May;45(2):187-195 [FREE Full text] [doi: [10.1016/j.rdc.2019.01.008](https://doi.org/10.1016/j.rdc.2019.01.008)] [Medline: [30952392](https://pubmed.ncbi.nlm.nih.gov/30952392/)]
15. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. *Annu Rev Public Health* 2019 Apr 01;40(1):127-146 [FREE Full text] [doi: [10.1146/annurev-publhealth-040218-044008](https://doi.org/10.1146/annurev-publhealth-040218-044008)] [Medline: [30601717](https://pubmed.ncbi.nlm.nih.gov/30601717/)]
16. Thorne S. *Interpretive Description Qualitative Research for Applied Practice*. Walnut Creek, California: Left Coast Press; 2016.
17. The complete guide to the wheel of life. The Coaching Tools Company. 2022 Jul 25. URL: <https://www.thecoachingtoolscompany.com/wheel-of-life-complete-guide-everything-you-need-to-know/> [accessed 2022-08-25]
18. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychologist* 2000;55(1):68-78. [doi: [10.1037/0003-066x.55.1.68](https://doi.org/10.1037/0003-066x.55.1.68)]
19. Oyserman D, Dawson A. Identity-based motivation and the paradox of the future self: getting going requires thinking about time (later) in time (now). *Behav Brain Sci* 2019 Dec 12;42:E269. [doi: [10.1017/s0140525x19000633](https://doi.org/10.1017/s0140525x19000633)]
20. Franklin M, Lewis S, Willis K, Bourke-Taylor H, Smith L. Patients' and healthcare professionals' perceptions of self-management support interactions: systematic review and qualitative synthesis. *Chronic Illn* 2018 Jun 22;14(2):79-103. [doi: [10.1177/1742395317710082](https://doi.org/10.1177/1742395317710082)] [Medline: [28530114](https://pubmed.ncbi.nlm.nih.gov/28530114/)]
21. Lee AA, Piette JD, Heisler M, Janevic MR, Rosland A. Diabetes self-management and glycemic control: the role of autonomy support from informal health supporters. *Health Psychol* 2019 Feb;38(2):122-132 [FREE Full text] [doi: [10.1037/hea0000710](https://doi.org/10.1037/hea0000710)] [Medline: [30652911](https://pubmed.ncbi.nlm.nih.gov/30652911/)]
22. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. *Int J Cardiol* 2019 Oct 01;292:280-282 [FREE Full text] [doi: [10.1016/j.ijcard.2019.05.066](https://doi.org/10.1016/j.ijcard.2019.05.066)] [Medline: [31171391](https://pubmed.ncbi.nlm.nih.gov/31171391/)]
23. Schroeder J, Karkar R, Murinova N, Fogarty J, Munson SA. Examining opportunities for goal-directed self-tracking to support chronic condition management. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2019 Dec 11;3(4):1-26 [FREE Full text] [doi: [10.1145/3369809](https://doi.org/10.1145/3369809)] [Medline: [32656490](https://pubmed.ncbi.nlm.nih.gov/32656490/)]
24. Hilty D, Chan S, Torous J, Luo J, Boland R. A framework for competencies for the use of mobile technologies in psychiatry and medicine: scoping review. *JMIR Mhealth Uhealth* 2020 Feb 21;8(2):e12229 [FREE Full text] [doi: [10.2196/12229](https://doi.org/10.2196/12229)] [Medline: [32130153](https://pubmed.ncbi.nlm.nih.gov/32130153/)]
25. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. *Telemed J E Health* 2015 Dec;21(12):1038-1041. [doi: [10.1089/tmj.2015.0002](https://doi.org/10.1089/tmj.2015.0002)] [Medline: [26171663](https://pubmed.ncbi.nlm.nih.gov/26171663/)]
26. Aria R, Archer N. The role of support and sustainability elements in the adoption of an online self-management support system for chronic illnesses. *J Biomed Inform* 2019 Jul;95:103215 [FREE Full text] [doi: [10.1016/j.jbi.2019.103215](https://doi.org/10.1016/j.jbi.2019.103215)] [Medline: [31132456](https://pubmed.ncbi.nlm.nih.gov/31132456/)]
27. Ma J, Yank V, Xiao L, Lavori PW, Wilson SR, Rosas LG, et al. Translating the Diabetes Prevention Program lifestyle intervention for weight loss into primary care: a randomized trial. *JAMA Intern Med* 2013 Jan 28;173(2):113-121 [FREE Full text] [doi: [10.1001/2013.jamainternmed.987](https://doi.org/10.1001/2013.jamainternmed.987)] [Medline: [23229846](https://pubmed.ncbi.nlm.nih.gov/23229846/)]
28. Kivelä K, Elo S, Kyngäs H, Kääriäinen M. The effects of nurse-led health coaching on health-related quality of life and clinical health outcomes among frequent attenders: a quasi-experimental study. *Patient Educ Couns* 2020 Aug;103(8):1554-1561. [doi: [10.1016/j.pec.2020.02.026](https://doi.org/10.1016/j.pec.2020.02.026)] [Medline: [32111383](https://pubmed.ncbi.nlm.nih.gov/32111383/)]
29. Denzin N. *The SAGE Handbook of Qualitative Research*. Thousand Oaks, California: SAGE Publications; 2011.
30. Sandelowski M. Triangles and crystals: on the geometry of qualitative research. *Res Nurs Health* 1995 Dec;18(6):569-574. [doi: [10.1002/nur.4770180612](https://doi.org/10.1002/nur.4770180612)] [Medline: [7480857](https://pubmed.ncbi.nlm.nih.gov/7480857/)]
31. Aureli F, Schino G. Social complexity from within: how individuals experience the structure and organization of their groups. *Behav Ecol Sociobiol* 2019 Jan 19;73(1):6. [doi: [10.1007/s00265-018-2604-5](https://doi.org/10.1007/s00265-018-2604-5)]
32. Jowett S. Creating a successful and effective coaching environment through interpersonal sports coaching. In: *Sport and Exercise Psychology: Practitioner Case Studies*. Hoboken, New Jersey, United States: Wiley; 2016.

## Abbreviations

- AS:** active surveillance
- ID:** interpretive description
- mHealth:** mobile health
- PCa:** prostate cancer
- SDT:** self-determination theory

*Edited by A Mavragani; submitted 25.10.21; peer-reviewed by C Dorfman, L Welch, C Memering, S De La Torre, E Baker; comments to author 27.05.22; revised version received 24.06.22; accepted 30.06.22; published 08.09.22.*

*Please cite as:*

*Obro LF, Osther PJS, Ammentorp J, Pihl GT, Heiselberg KK, Krogh PG, Handberg C*

*An Intervention Offering Self-management Support Through mHealth and Health Coaching to Patients With Prostate Cancer: Interpretive Description of Patients' Experiences and Perspectives*

*JMIR Form Res 2022;6(9):e34471*

*URL: <https://formative.jmir.org/2022/9/e34471>*

*doi: [10.2196/34471](https://doi.org/10.2196/34471)*

*PMID: [35925751](https://pubmed.ncbi.nlm.nih.gov/35925751/)*

©Louise Faurholt Obro, Palle Jörn Sloth Osther, Jette Ammentorp, Gitte Thybo Pihl, Kasper Kvals Heiselberg, Peter Gall Krogh, Charlotte Handberg. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 08.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Understanding Gender Biases and Differences in Web-Based Reviews of Sanctioned Physicians Through a Machine Learning Approach: Mixed Methods Study

Julia Barnett<sup>1</sup>, BBA, MS; Margrét Vilborg Bjarnadóttir<sup>2</sup>, BSc, PhD; David Anderson<sup>3</sup>, BS, PhD; Chong Chen<sup>2</sup>, BS, MS

<sup>1</sup>Technology and Social Behavior, Northwestern University, Evanston, IL, United States

<sup>2</sup>Robert H. Smith School of Business, University of Maryland, College Park, MD, United States

<sup>3</sup>Villanova School of Business, Villanova University, Villanova, PA, United States

**Corresponding Author:**

Julia Barnett, BBA, MS

Technology and Social Behavior

Northwestern University

2240 Campus Drive

Room 2-168 (Attention: Dept of Comm Studies)

Evanston, IL, 60208

United States

Phone: 1 847 491 7530

Email: [JuliaBarnett@u.northwestern.edu](mailto:JuliaBarnett@u.northwestern.edu)

## Abstract

**Background:** Previous studies have highlighted gender differences in web-based physician reviews; however, so far, no study has linked web-based ratings with quality of care.

**Objective:** We compared a consumer-generated measure of physician quality (web-based ratings) with a clinical quality outcome (sanctions for malpractice or improper behavior) to understand how patients' perceptions and evaluations of physicians differ based on the physician's gender.

**Methods:** We used data from a large web-based physician review website and the Federation of State Medical Boards. We implemented paragraph vector methods to identify words that are specific to and indicative of separate groups of physicians. Then, we enriched these findings by using the National Research Council Canada word-emotion association lexicon to assign emotional scores to reviews for different subpopulations according to gender, gender and sanction, and gender and rating.

**Results:** We found statistically significant differences in the sentiment and emotion of reviews between male and female physicians. Numerical ratings are lower and sentiment in text reviews is more negative for women who will be sanctioned than for men who will be sanctioned; sanctioned male physicians are still associated with positive reviews.

**Conclusions:** Given the growing impact of web-based reviews on demand for physician services, understanding the different dynamics of reviews for male and female physicians is important for consumers and platform architects who may revisit their platform design.

(*JMIR Form Res* 2022;6(9):e34902) doi:[10.2196/34902](https://doi.org/10.2196/34902)

## KEYWORDS

gender; natural language processing; web-based reviews; physician ratings by customer; text mining

## Introduction

### Background

Web-based reviews of physicians play an important role in patients' searches for providers. A 2017 National Institutes of Health survey found that 39% of adults used web-based reviews

to help them decide which physician to see [1]. Although ratings are widely used, they are also sparse, with many physicians having only 1 or 2 ratings on any given site. In addition, up to 90% of all web-based reviews are positive [2]. Positive reviews have been estimated to increase physician demand by as much as 7% [3]. Thus, to fully understand the impact of web-based

reviews on individual providers and the health care system as a whole, it is important to identify when and why physicians are given negative reviews and whether those negative reviews actually reflect reality.

There are many reasons to believe that web-based reviews may be unhelpful. For example, they are subject to review fraud, as Hu et al [4] estimated that 10% of all web-based reviews are fraudulent. In addition, they are typically nonexpert reviews of expert services; that is, credence services, and customer reviews of credence services are generally thought to be unhelpful [5]. For both reasons, Dr Peter Carmel, president of the American Medical Association, has argued that “anonymous online opinions of physicians should be taken with grain of salt and should not be a patient’s sole source of information when looking for a new physician” [6]. Studies have further found that although web-based reviews matched peer evaluations of physicians in outpatient specialties, this was not the case for inpatient surgical specialties [7].

Gender adds another layer of complexity to questions related to the influence, validity, and value of web-based reviews. Studies have shown that female physicians are given more negative reviews and that women are rated as less amicable [8]. Although previous studies have shown differences in review content based on physician gender, our study adds a critical dimension by considering an external indicator of physician quality. We used data from the Federation of State Medical Boards on physicians who have been sanctioned by their state medical boards for unsuitability to practice medicine, either for negligence, malpractice, or other improper behavior. Sanctions range from probation to complete revocation of the offending physician’s medical license. As receiving a sanction is an objective marker of low-quality medical care, at least for some physicians, looking at sanctions gives us a way to quantify physician quality, which is a notoriously difficult task. We showed that women receive systematically different reviews from men and that female physicians who will be sanctioned in the future are rated lower and receive more negative comments in their reviews than similarly situated male physicians.

## Related Literature

### *Physician’s Gender*

Previous studies have explored gender differences in light of how physicians consult and communicate with their patients. Studies have concluded that female physicians are generally more communicative and interpersonal than male physicians as they focus more on building partnership, asking questions, and providing information, which results in long medical appointments with female physicians [9]. This long consultation duration reduces the volume of consultations that female physicians can provide [10]. Some studies have explored the reasons for long consultations and the role of gender in medical decision-making. For instance, when diagnosing coronary heart disease, female physicians are more engaged with the historical presentation of the patient’s condition and more likely to be affected by the patient’s gender than male physicians. Greenwood et al [11] showed that female patients who had heart

attacks treated by male physicians had significantly higher mortality than female patients treated by female physicians.

Web-based physician reviews have been studied from diverse angles. We divided our summary of the literature into review content, how reviews correlate with peer ratings, fraudulent reviews, sentiment analysis, and finally, the impact of physician reviews.

### *Content of Physician Reviews*

Hao and Zhang [12] implemented latent Dirichlet allocation (LDA) as a topic modeling technique for textual review of data about Chinese physicians in 4 major specialty areas and identified popular review topics, including professionalism and showing appreciation for physicians’ detailed symptom descriptions [12]. A second study identified patient satisfaction, staff, and access as important themes in the reviews studied [13].

An extensive analysis of reviews from US health care review websites by Thawani et al [14] found that female physicians receive lower ratings overall, even after accounting for specialty. Comments about female physicians are more likely to be related to their interpersonal skills, whereas for male physicians, comments focus more on professionalism and helpfulness. Marrero et al [15] further examined a subset of the same data to understand the influence of gender on how patients both perceive and evaluate their surgeons, confirming that women are evaluated more positively for social interactions and men for technical aptitude.

### *Validity of Physician Reviews*

McGrath et al [7] examined the validity of patient-generated web-based physician reviews and found that validity is affected by physician specialty. For specialties such as family medicine, allergies, internal medicine, and pediatrics, the web-based ratings of physicians listed as a *top doctor* by their peers are significantly higher than the ratings of those without this peer-generated quality indicator. Kordzadeh [16] showed that the ratings listed on hospital websites are systematically higher than those on outside commercial physician rating sites such as RateMDs and Google Reviews.

### *Sentiment of Physician Reviews*

Wallace et al [17] developed a factorial LDA model to jointly identify both sentiment and topics from reviews. By incorporating the factorial LDA output into regression analysis, they further found that positive sentiment is associated with health care measurements such as patients’ revisit probability and health care costs and that the model can explain more variance than models using only rating information. Similarly, Rivas et al [18] developed a dependency tree-based classifier to capture patterns from each review, which can be used to sort physician reviews into a 2D classification system based on topic and polarity. Waltena et al [19] focused on the impact of sentiment on topic extraction in hospital reviews, and by adding 2 topics representing positive and negative sentiment in latent semantic analysis, the authors successfully reduced the bias owing to sentiment on the subjects of topics.



## Impact of Physician Reviews

The impact of web-based physician reviews on patient choice remains an active area of research. Xu et al [3] explored the interaction between web-based physician reviews and physician demand and concluded that the number of reviews and disclosure of reviewer identity are positively related to physician demand but negatively correlated with review length. Through a counterfactual experiment, they found that strategies for improving ratings (eg, disclosing reviewers' identities and limiting review length) can increase the demand for a physician by as much as 7.24%. However, improving the operational process or platform design can increase physician demand even further. Li et al [20] studied how web-based reviews and physicians' gender affect patients' primary care physician choices. The results indicated that among physicians whose skills are endorsed in reviews, if a female physician is endorsed for their interpersonal characteristics, such as compassion and personableness, they are more likely to be chosen than a male physician endorsed for the same reasons. However, this kind of gender effect is not observed among physicians endorsed for their technical skills. Bedside manner, diagnosis accuracy, patients' waiting time, and consultation length are critical in patients' choice of a physician [3].

Our study is the first to analyze the content of patient reviews of physicians across genders using natural language processing tools that accounts for differences in ratings and sanction status. This allowed us to understand both the set of criteria on which male and female physicians are evaluated and the impact of poor performance (as measured by sanctions). We further applied an emotional index to understand, in a multidimensional way, the tones of the different types of reviews based on ratings, gender, and sanction status. More specifically, in this study, we aimed to determine whether reviews of women systematically differ from those of men. In particular, we aimed to discover whether female physicians are rated lower at baseline than male physicians and whether female physicians experience larger reputational penalties than male physicians for low-quality services (as indicated by sanctions from the state medical board).

## Methods

### Data

Our data were collected from 2 sources: physician reviews were obtained from RateMDs and combined with physician sanction data from the Federation of State Medical Boards [21].

The data from RateMDs include physicians' average ratings on a 1- to 5-star scale. Reviewers rate the overall experience and 4 other defined categories: helpfulness, knowledgeability, punctuality, and staff. The data further contain the text of the reviews.

State licensing boards issue sanctions to physicians for issues related to their suitability to practice medicine in each state. Reasons for sanctions include, but are not limited to, serious malpractice, performing unnecessary treatment, fraudulent billing, and abuse of patients. We collected every review posted between October 2004 and August 2011 and matched it by name, location, state, and specialty with the database of

sanctioned physicians from the Federation of State Medical Boards. We removed any reviews of physicians that were made after they were sanctioned, so any official sanction does not affect the content of the reviews. In total, we obtained 403,470 reviews of 134,973 physicians across the United States. In our data, men were more than twice as likely as women to be sanctioned; 1.7% (1629/95,831) of all male physicians were sanctioned, whereas only 0.64% (250/39,142) of female physicians were sanctioned.

The web-based reviews from RateMDs were merged with the state medical board sanction data by matching physician name (including matching using a dictionary of common nicknames [eg, Kate for Katherine]), state, specialty, medical school, and graduation year (where available). The physicians in the sanction data who we could not perfectly match owing to multiple matches or no matches (and which amounted to <5% of the sample) were excluded from the study.

## Methodology

### Overview

The field of text mining and natural language processing is growing rapidly, with many emerging techniques available to analyze text and discover patterns in documents via automated procedures. In their book, *Foundations of Statistical Natural Language Processing*, Manning and Schütze [22] stated that the availability of large text corpora has changed the scientific approach to language in linguistics and cognitive science. Therefore, phenomena that were previously undetectable or seemingly uninteresting have become the central focus of lexical analysis. Taking advantage of some of these new developments, in this study, we implemented paragraph vector (as described in the following sections) and used a word-emotion association lexicon on the corpus of physician reviews to analyze the data in a nuanced manner.

### Data Preprocessing

To make the raw data analyzable, we performed a series of tasks. First, the reviews were converted to lowercase, so that capital letters are treated the same as lowercase letters. Second, punctuation was removed because it typically adds unnecessary noise to word models. Third, *stopwords*, defined as unimportant words that are overly common (eg, "the," "and," and "is") were removed using a freely available System for the Mechanical Analysis and Retrieval of Text stopword list built by Salton and Buckley and sourced from web-based Appendix 11 of the paper by Lewis et al [23]. Fourth, we removed numbers because, similar to punctuation, they add noise to the analysis. On the remaining words in the corpus, we performed *stemming* using the Porter stemming algorithm [24,25]. Stemming is the act of reducing words to their root form (eg, "practice," "practicing," and "practiced" become "practic"). This allows models to treat these words as one concept rather than as separate ideas. As we had a limited-sized data set, we applied all the preprocessing steps mentioned previously to maximize insights from a concise vocabulary. Although the removal of stopwords resulted in some locally unnatural word sequences (such as articles not appearing before nouns), we found that this did not hinder our analysis.

### Analytical Techniques

In this study, we applied a paragraph vector framework [26], a natural language processing method that represents each word or document as a dense vector (ie, a location in  $\mathbb{R}^d$  space), called an embedding, which is then used as an input to train a model to predict co-occurrence of words. We used the paragraph vector distributed bag-of-words model, which uses words from a given width window to predict the next word in the document. In this framework, “kind” is located closer to “nice” than “surgery” because “nice” has a much higher probability than “surgery” of being found in similar contexts as “kind.” We used the paragraph vector model to generate an embedding of words, which can be used to calculate the similarity (via cosine similarity) between any set of words or documents. Henceforth, we refer to the cosine similarity between words or documents as the *similarity score*.

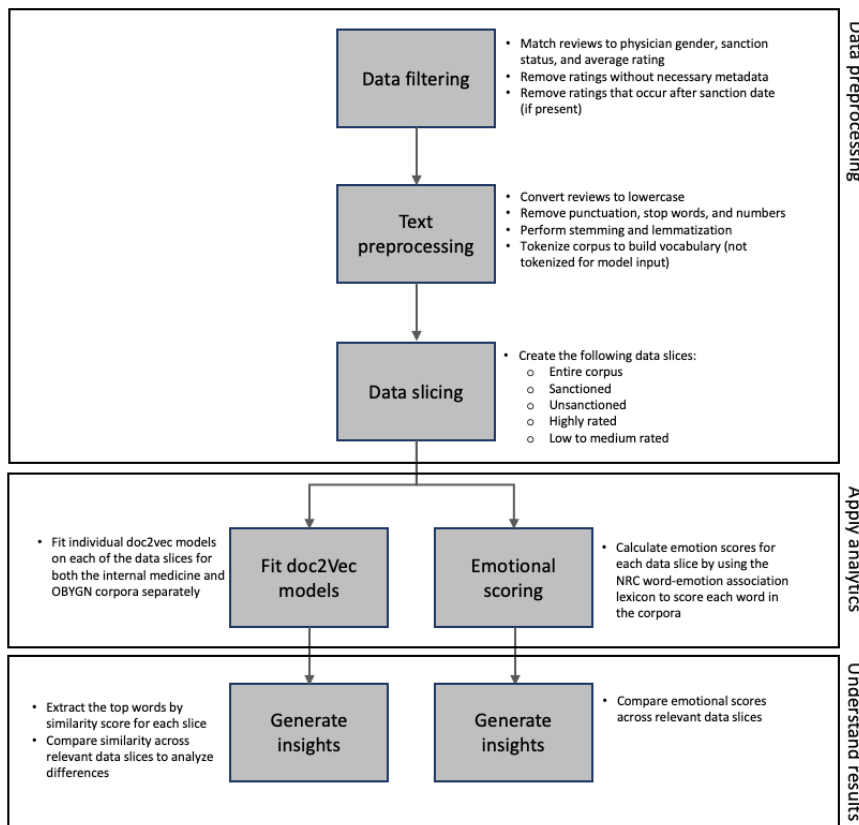
For each data slice (eg, sanctioned physicians), we trained a paragraph vector model. Once the model was trained, we could use the embedding to identify words associated with the medical reviews of different types of physicians (eg, based on gender). To compare specific differences across a physician population, we concatenated every review from one specific subset of data (eg, sanctioned *male* physicians) and found the similarity scores of this document with each word within the corpus. Then, we repeated this process for the complementary subset (eg, sanctioned *female* physicians) and compared the similarity scores for each subset. We extracted the words with the largest

differences between the subsets. For example, we computed the similarity of *wait* to the female corpus of reviews, computed the similarity of *wait* to the male corpus of reviews, and calculated the difference in these scores. Our analysis focused on the words with the highest absolute difference between the similarity scores for one subset of reviews (typically female) and the complementary subset (typically male).

To understand the *emotional* nature of the reviews, we used the NRC word-emotion association lexicon [27] to attribute sentiment and emotional scores to the corpus (NRC stands for the National Research Council Canada, but the lexicon is commonly referred to as the NRC emotion lexicon). This lexicon created an *afinn* dictionary by rating words on a scale of 8 emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Using the scores from this lexicon, we were able to both rate reviews on an aggregate emotional scale (how *emotional* the document is as a whole) and rank them for each of the 8 emotions. More specifically, for each data cut (eg, sanctioned female physicians), each word in each physician’s review was scored based on the emotional score of the word, and then, average physician score was derived by averaging all physicians’ emotional scores. Understanding these emotional scores allowed us to develop a deep understanding of the criteria that patients use to evaluate female and male physicians and how those criteria differ.

We have summarized the methodological approach in [Figure 1](#).

**Figure 1.** Analysis flowchart. NRC: National Research Council Canada; OBGYN: obstetrics and gynecology.



## Implementation and Hypertuning

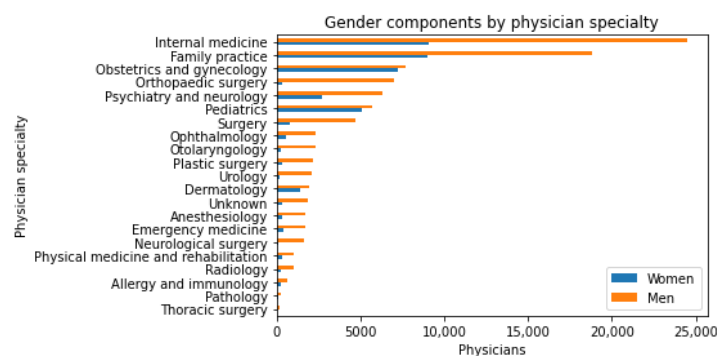
We used doc2vec [28], a Python package implementing paragraph vectors, to learn how patients review physicians differently across gender, sanctions, and ratings (both in isolation and interaction). We performed a standard doc2vec implementation to learn the paragraph vectors of the following:

1. Gender (male and female)
2. Composite label of gender and sanction (female sanctioned, female unsanctioned, male sanctioned, and male unsanctioned)
3. Composite label of gender and rating (female high rating, female low to medium rating, male high rating, and male low to medium rating), where we defined high rating as  $\geq 4$  stars and low to medium rating as  $< 4$  stars

To overcome majority bias, we sampled an equal number of reviews for each group. We trained the models independently for the different metadata cuts, rather than treating each separate review as an individual document. By fitting the different groups separately, we were able to understand the specific lexicons associated with each metadata cut (gender, sanction, and rating). Then, we analyzed the similarity scores of words to their respective corpora and compared the scores.

We pretrained the paragraph vector framework, using the continuous bag-of-words algorithm to tune the hyperparameters, by testing the most similar words to several words such as “knowledgeable,” “wonderful,” “caring,” and “rude.” We ran multiple variations of the model to identify the best settings. The results were consistent across different parameters, which gave us confidence in the robustness of the final model. We

**Figure 2.** Number of physicians in each specialty, broken down by gender.



**Table 1** highlights the average star ratings overall and for each of the main categories present in the reviews (helpfulness, knowledgeability, punctuality, and staff). This pattern is consistent across specialties including internal medicine and OBGYN. Furthermore, unsanctioned physicians receive higher ratings than sanctioned physicians. We note that the *staff* category ratings may not be reflective of the physician’s medical capabilities. In all cases, the average rating for men is higher than that for women. These differences are statistically significant ( $P < .001$ , evaluated with 2-tailed  $t$  tests) both when comparing unsanctioned male physicians with unsanctioned female physicians and when comparing unsanctioned female or male physicians with sanctioned female or male physicians.

have listed the exact parameter settings in [Multimedia Appendix 1](#) and summarized the results of the paragraph vector model for “knowledgeable,” “wonderful,” “caring,” and “rude” in [Figure S2 in Multimedia Appendix 1](#).

## Ethical Considerations

All the data used in this study are publicly available and do not contain identifiable private information about individuals. Thus, this study was not deemed to require institutional review board review. After merging sanction data with review data by name, specific physician identities were removed from the data set and not used in the analyses.

## Results

### Data Overview

**Figure 2** shows the number of physicians in each specialty by gender. Internal medicine and family practice are the 2 most common specialties in our data. The figure highlights that there are more male physicians than female physicians in every specialty; overall, 29% (39,142/134,973) of the physicians in the sample are women. This gender imbalance is easily noticeable in the more common disciplines; internal medicine has the highest number of female physicians, but there are still twice as many male physicians. The imbalance is even more prominent in some small disciplines such as orthopedic surgery and neurological surgery, where men outnumber women 23:1 and 15:1, respectively. Obstetrics and gynecology (OBGYN) and pediatrics departments are more balanced in terms of gender, with an approximately even ratio of men to women.

When the ratings of sanctioned male physicians are compared with those of sanctioned female physicians, the absolute differences are of similar magnitude; however, owing to the small size of the sanctioned population, the differences are not statistically significant. Among sanctioned physicians, female physicians receive lower ratings (by an average of approximately 0.1 stars) than male physicians (not considering specialties). The difference between genders among sanctioned physicians is greater than that among unsanctioned physicians, especially for those rated around average for helpfulness and knowledgeability. A detailed breakdown of the number of sanctioned physicians is provided in [Table S1 in Multimedia Appendix 1](#).

**Table 1.** Average star rating (out of 5 stars) overall and for the 4 RateMDs score categories for the whole sample of physicians. Ratings are separated by gender and sanction status.

Categories	Full sample (n=134,973), mean (SD)	Female physicians (n=39,142, 29%), mean (SD)		Male physicians (n=95,831, 71%), mean (SD)	
		Unsanctioned (n=38,892, 99.36%)	Sanctioned (n=250, 0.64%)	Unsanctioned (n=94,202, 98.30%)	Sanctioned (n=1629, 1.69%)
Overall	3.86 (1.12)	3.81 (1.12)	3.41 (1.21)	3.89 (1.12)	3.52 (1.24)
Helpfulness	3.89 (1.36)	3.85 (1.35)	3.47 (1.44)	3.90 (1.36)	3.59 (1.48)
Knowledgeability	4.03 (1.25)	3.99 (1.24)	3.62 (1.35)	4.06 (1.25)	3.75 (1.38)
Punctuality	3.83 (1.18)	3.77 (1.18)	3.39 (1.27)	3.86 (1.18)	3.43 (1.31)
Staff	3.67 (1.3)	3.61 (1.3)	3.02 (1.41)	3.70 (1.3)	3.19 (1.45)

On average, each physician receives 3 reviews, with an average length of 55.7 (SD 47.65) words. In general, lower-ranked physicians receive longer reviews than higher-ranked physicians; people have more to say about an experience they are dissatisfied with. On average, women receive longer reviews than men. An exception is in the OBGYN field—in this specialty, patients have more to say about a sanctioned male physician than they do about a sanctioned female physician.

However, on the whole and across specialties, the group with the longest reviews are the low to medium-ranked women. Given what we know from our subsequent content analysis, this is because patients have longer and more negative comments to make about women, whereas reviews of male physicians are short and more positive. As highlighted in [Table 2](#), these trends hold when we break down the length analysis by specialty.

**Table 2.** Average review length for sanctioned and unsanctioned male physicians and female physicians in all specialties, internal medicine, and OBGYN<sup>a</sup>, measured in number of words.

Categories	All specialties (n=134,973)		Internal medicine (n=33,549)		OBGYN (n=15,001)	
	Female (n=39,142, 29%), mean (SD)	Male (n=95,831, 71%), mean (SD)	Female (n=9087, 27.09%), mean (SD)	Male (n=24,462, 72.91%), mean (SD)	Female (n=7268, 48.45%), mean (SD)	Male (n=7733, 51.55%), mean (SD)
Overall	50.1 (36.4)	45.7 (36.1)	45.8 (36.3)	41.5 (35.2)	58 (34.6)	54.5 (34.9)
Sanctioned	48.7 (36.6)	47.6 (37.1)	55.1 (34)	42.9 (35.8)	46 (30.8)	51.2 (38.7)
Unsanctioned	50.2 (36.4)	45.6 (36.1)	45.7 (36.3)	41.4 (35.2)	58.2 (34.7)	54.5 (34.9)
High rating	39 (31)	37.2 (31.3)	35.9 (31)	33.4 (29.8)	45.4 (29.6)	47.8 (32)
Low to medium rating	61.6 (38)	56.3 (38.8)	57.3 (38.5)	53 (38.9)	69.7 (34.8)	65 (36.7)

<sup>a</sup>OBGYN: obstetrics and gynecology.

The nature of physicians' work differs between specialties, which in turn may influence web-based reviews. Therefore, to remove the impact of specialty, our analysis in this study focused on internal medicine (the most common type of physician reviewed). In addition, we conducted the analysis on OBGYN reviews and compared the results with those for internal medicine (detailed results for the OBGYN reviews are available in [Multimedia Appendix 1](#)). This allowed us to compare results across medical specialties, but the OBGYN results are particularly interesting, as we can be confident that most reviews are written by women, giving us further insight into the differences in results. Following these analyses, we compared the length and emotion of the reviews.

We examined the differences between gender and reviews in 3 ways: first, we analyzed male and female physicians; second, we studied both gender and rating; and third, we analyzed the interaction of sanction and gender. For each of these analyses, a separate doc2vec model was trained on the relevant corpus (eg, the entire corpus, reviews of sanctioned physicians, or

highly ranked reviews). For each analysis (eg, male physician vs female physician in highly ranked reviews), we extracted the top words by similarity score to the paragraph vector of concatenated female reviews and concatenated male reviews, respectively, in the relevant subset of data, and then compared the differences. Our analysis focused on the words with the greatest absolute difference between the similarity scores for female and male reviews. Additional and complementary results are available in [Multimedia Appendix 1](#).

### Review Comparison Between Genders

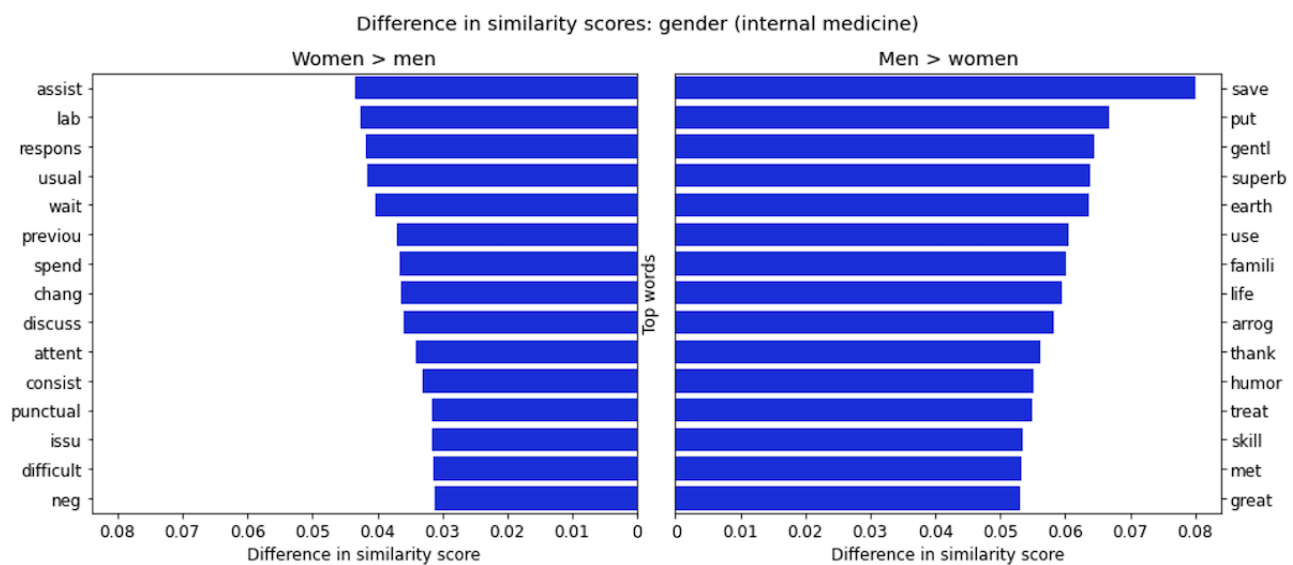
To examine the relative similarity scores of the words used in the corpus to describe men and women, we extracted the top words by similarity score (omitting procedural-type words, eg, "appt" and "said" for analysis purposes) for the subset of male physician reviews and female physician reviews, as summarized in [Figure 3](#). This figure presents the top 15 words with the largest difference between similarity scores to the document vector for concatenated female reviews and concatenated male

reviews, with the left pane showing the 15 words that scored highest for the female reviews and the right pane showing those that scored highest for male reviews.

This led us to several interesting observations. For example, “assist,” “neg,” “difficult,” “wait,” “punctual,” and “issue” were scored as more similar to female physicians’ reviews than male physicians’ reviews. In contrast, “superb,” “gentl,” “famil,” “skill,” “humor,” and “great” were scored as more similar to male physicians’ reviews. The key takeaway is that even before we incorporate sanction and ranking data, we see stark differences between the ways male and female physicians are

evaluated, which supports the findings of previous analyses [14-16]. From this general comparison, without considering rating, we see that women are more often evaluated with a focus on punctuality, whereas men are much more likely to be praised for their technical abilities and bedside manner. To confirm that these underlying frequencies are statistically significant, we performed a chi-square test for the top words presented in Figure 3 (the null hypothesis was that there is no difference in these frequencies, and the alternative hypothesis was that there are differences in these frequencies not equal to 0) and found all of them to be significant ( $P < .001$ ).

**Figure 3.** Difference in similarity scores for top words in reviews of male and female internal medicine physicians. The x-axis represents the absolute difference in similarity score for the given words to the document vector of concatenated reviews for all women and all men. The figure displays the top 15 words; the biggest differences in similarity scores are for the female subset of reviews over male reviews (left pane) and the male subset of reviews over female reviews (right pane).



### Review Comparison Between Gender and Rating

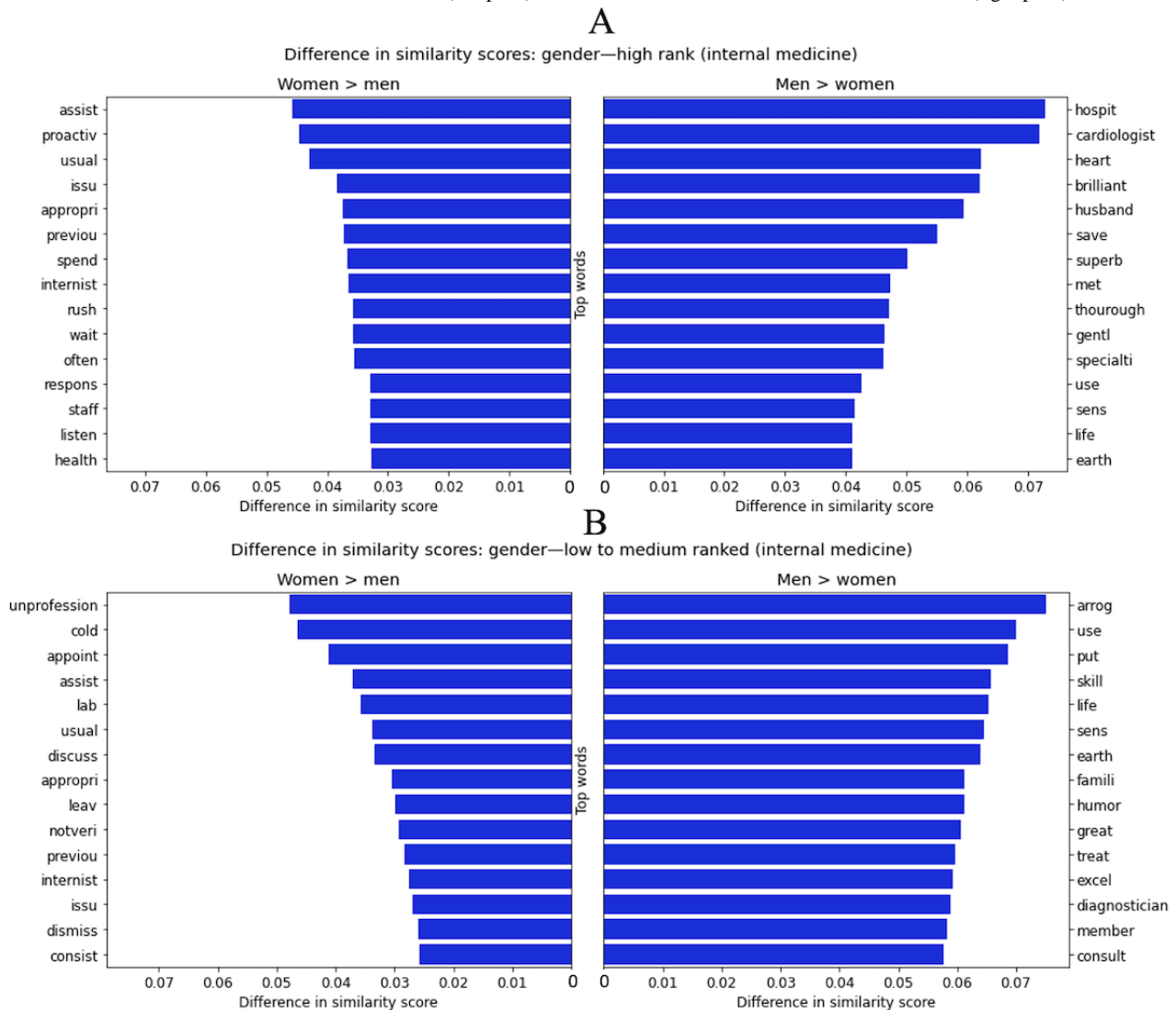
We used the approximate mean as the standard criterion for determining whether a rating was *high* ( $>4$  stars) or *low to medium* ( $\leq 4$  stars). Then, we created document vectors for the following subsets of concatenated reviews: (1) reviews rating female physicians highly, (2) reviews rating male physicians highly, (3) reviews rating female physicians as medium to low, and (4) reviews rating male physicians as medium to low. We repeated this analysis, focusing first on high-ranked female and male physicians and second on low to medium-ranked female and male physicians. The results are shown in Figure 4. We again compared the top words by absolute difference in similarity score between men and women within the high reviews first and then within the low to medium reviews.

For highly ranked physicians, the words that are the most associated with female physicians’ reviews over the corpora or male physicians’ reviews tend to either describe the timeliness of the visit (eg, “wait” and “rush”), liken female physicians to workers in supporting roles, or evaluate staff in those supporting

roles (eg, “assist” and “staff”). In contrast, the corpora of male physicians’ reviews are more likely to contain words that are medically technical (eg, “hospit,” “cardiologist,” “skill,” or “diagnostician”) or simply glowing endorsements (eg, “brilliant,” “superb,” and “greatest”). These findings are summarized in Figure 4A.

Despite these discrepancies, we note that highly ranked physicians generally garner positive text reviews regardless of gender. Gender differences become much more pronounced when focusing on low-ranked physicians. As summarized in Figure 4B, the words with the highest similarity scores for reviews for low to medium-ranked women are objectively much more negative (eg, “unprofession,” “cold,” “issu,” “dismiss,” and “notveri”) compared with the reviews of low to medium-ranked men (eg, “skill,” “sens,” “famili,” “humor,” “great,” and “excel”). The only objectively negative word that is much more likely to occur in these male physicians’ reviews is “arrog” for arrogance (a quality more often attributed to men than to women).

**Figure 4.** Difference in similarity scores for top words for (A) high-ranked and (B) low to medium-ranked men and women in internal medicine. The x-axis represents the absolute difference in similarity score for the given words to the document vector of concatenated reviews for all (A) high-ranked women and all men and (B) low to medium-ranked women and all men. The figure displays the top 15 words with the biggest differences in similarity scores for the female subset of reviews over male reviews (left pane) and the male subset of reviews over female reviews (right pane).

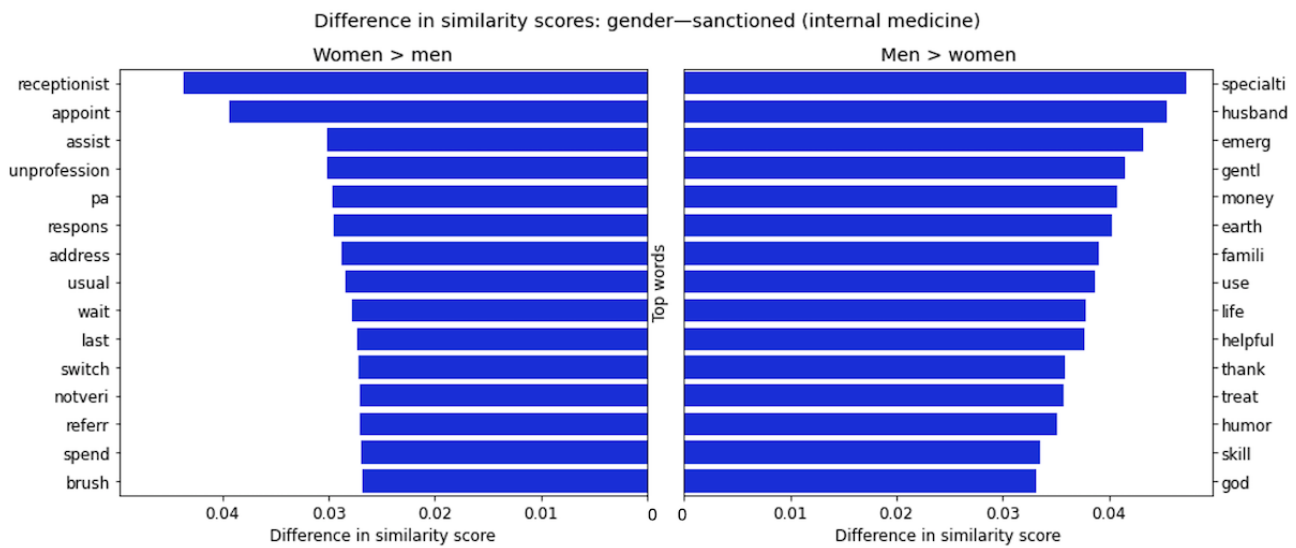


**Review Comparison Between Gender and Sanction**

As discussed previously, male physicians receive high ratings on average, but at the same time are more likely to be sanctioned. This motivated our independent analysis of reviews of sanctioned and unsanctioned physicians by gender. Owing to the low overall probability of sanctions (1879/134,973, 1.39% of our sample), the reviews of unsanctioned physicians mirror the general discrepancies between men and women. In contrast, the analysis of sanctioned physicians’ reviews reveals stark gender differences, as highlighted in Figure 5. The words with the highest probability of appearing in sanctioned women’s reviews have much more negative connotations than those in

sanctioned men’s reviews, whereas it is much more difficult to tell the difference between a sanctioned man and an unsanctioned man. Some of the words most highly associated with sanctioned male physicians are “specialti,” “gentl,” “helpful,” “thank,” “skill,” and “god,” whereas some of the highest scored words for sanctioned female physicians are “receptionist,” “unprofession,” “pa,” “wait,” and “notveri.” Words that are exclusive to the sanctioned male lexicon include “cardiologist,” “save,” “heart,” “hospit,” “superb,” “pleasur,” and “compassion,” which highlight the stark discrepancies even further because these words do not appear even once in a sanctioned female physician’s review (additional details are available in Figures S4 and S5 in Multimedia Appendix 1).

**Figure 5.** Difference in similarity scores for top words for sanctioned men and women in internal medicine. The x-axis represents the absolute difference in similarity score for the given words to the document vector of concatenated reviews for all sanctioned women and all sanctioned men. The figure displays the top 15 words with the biggest differences in similarity scores for the female subset of reviews over male reviews (left pane) and the male subset of reviews over female reviews (right pane).



**Emotion Scoring**

We analyzed the emotional scores of the reviews, generating an emotional score for each subset of physicians. We used the percentage of each top word appearing in each cut of the review corpus multiplied by the emotional score, repeated the process for each word in the lexicon to obtain a total score for each cut (eg, male, female, sanctioned women, and highly rated men), and then summed these scores within each subset. The emotion analysis is the only portion of this study in which we found noticeable differences between the 2 specialties analyzed—internal medicine and OBGYN. Therefore, we have included the results for both specialties in the main text.

In the plots below, the emotions are categorized as positive, negative, or neutral and listed alphabetically within each category in the following order: joy, positive, trust, anticipation, surprise, anger, disgust, fear, negative, and sadness.

First, we examined the differences in emotional scores between high-ranked and low to medium-ranked female physicians (Figure 6A) and between high-ranked and low to medium-ranked male physicians (Figure 6B). As expected, more positive emotions are much more likely to be found in high ratings of both men and women, with only small differences between men and women in both specialties analyzed.

When using gender (rather than rating) as the main dimension of analysis, we found that for internal medicine, reviews of men are much more emotional than those of women, for both positive and negative emotions, as demonstrated in Figure 7. A notable exception is that women’s reviews scored high on negative emotion. For OBGYN physicians (reviews that we can safely assume to be written mostly by women), the reviews are much more positive for men (overindexing on joy, positive, and trust), and the reviews of female physicians score notably high on anticipation, disgust, negativity, and sadness.

Next, we divided the analysis by gender and specialty, and then focused on the difference between sanctioned and unsanctioned physicians. The results are highlighted in Figures 8A and 8B. For female internal medicine physicians, the results are consistent with expectations; unsanctioned physicians score high on positive emotions, whereas sanctioned physicians score high on neutral and negative emotions. In contrast, for male internists, unsanctioned physicians score high across the emotional scale (however, the differences are generally small). The pattern for OBGYN physicians is very different—among female OBGYN physicians, there is great variability in the emotional scores, whereas among male OBGYN physicians, unsanctioned physicians score high on positive and neutral emotions, with very little difference in emotional scores on negative emotions.

**Figure 6.** Emotional score ratings for (A) female physicians' and (B) male physicians' reviews. The 10 emotions on the y-axis are categorized as positive, neutral, or negative (and arranged alphabetically within these categories). The x-axis plots the difference in the emotional score between the different groups. Positive numbers mean that an emotion scored high for high-ranked physicians, and negative numbers mean the emotion scored high for low to medium-ranked physicians. OBGYN: obstetrics and gynecology.

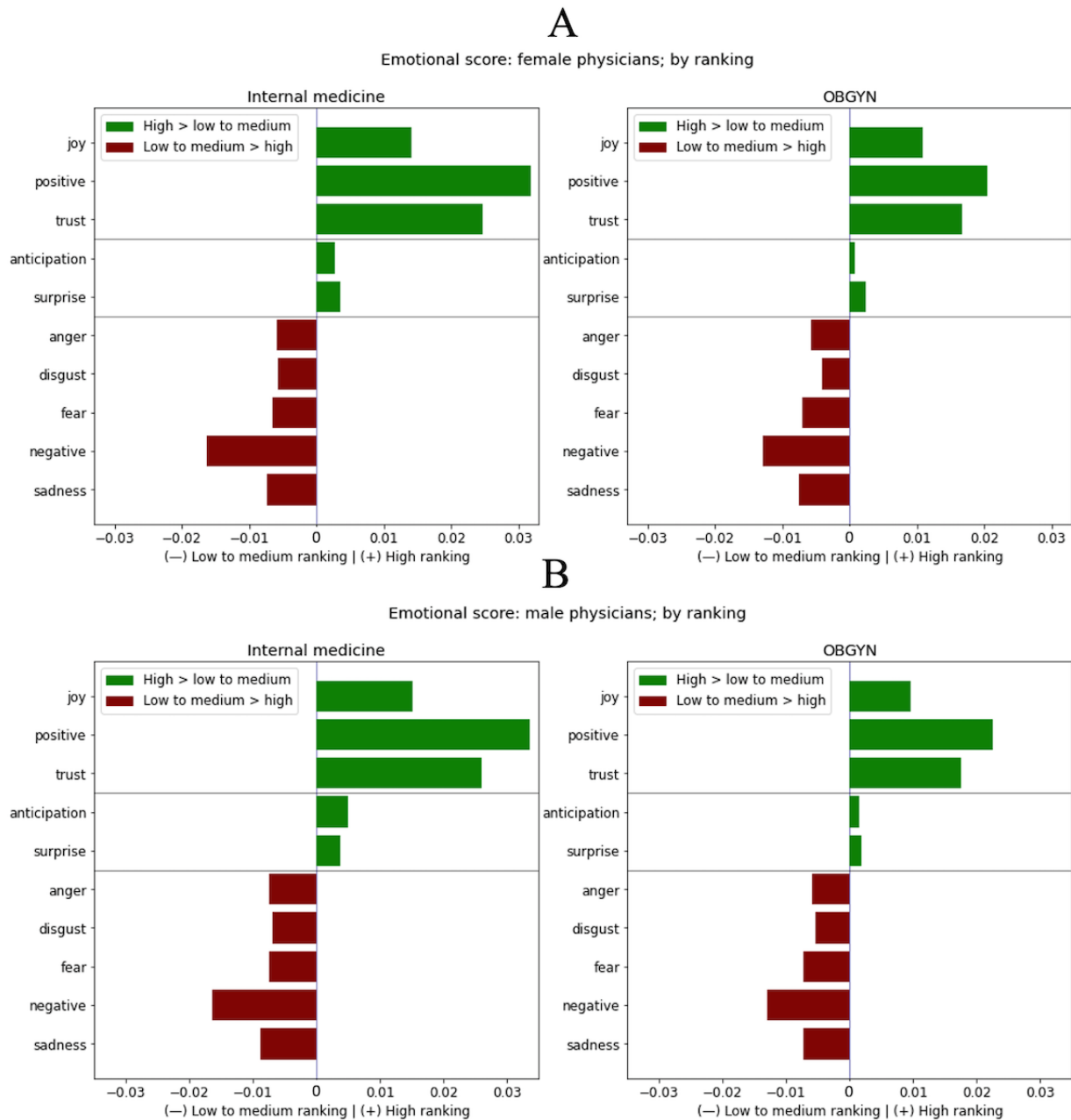




Figure 7. Emotional scores by gender for internal medicine and obstetrics and gynecology (OBGYN).

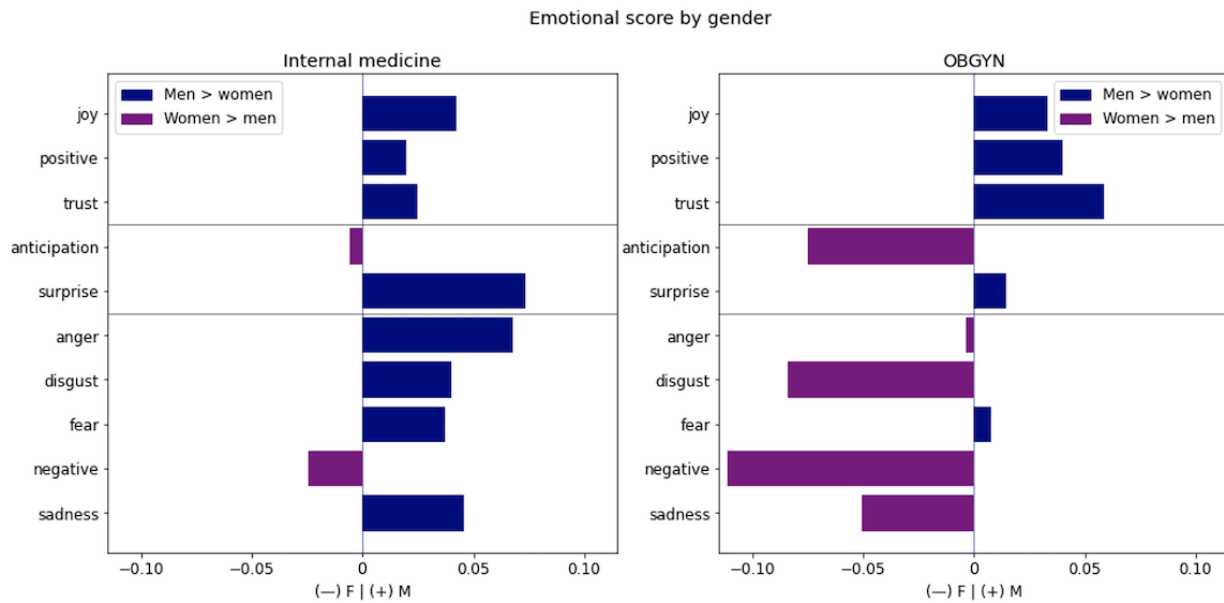
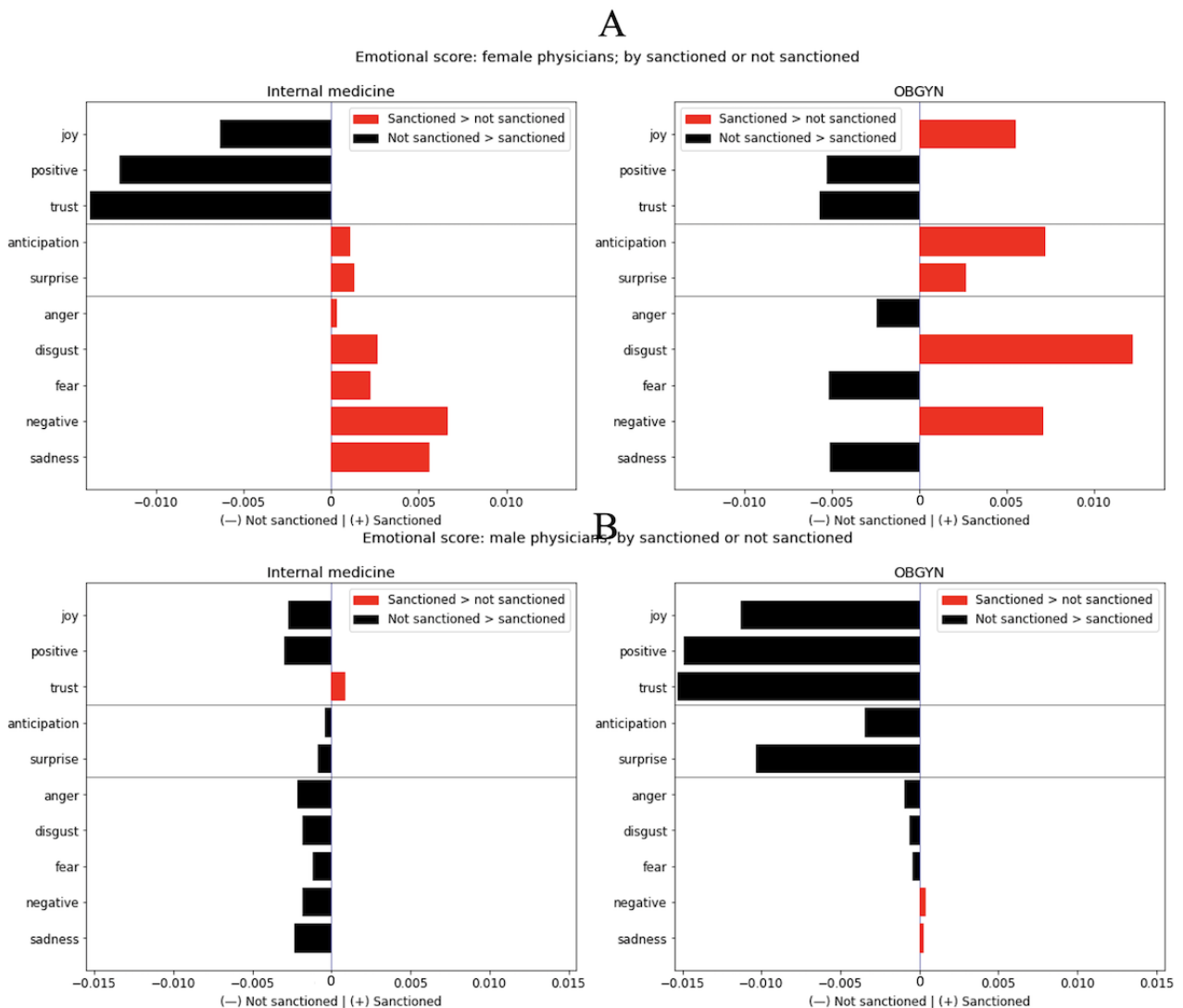


Figure 8. Emotional scores by sanction status for (A) female physicians and (B) male physicians for both internal medicine and obstetrics and gynecology (OBGYN).



Finally, we focused on the differences between sanctioned men and sanctioned women and summarize the results in Figure S18 in [Multimedia Appendix 1](#). In contrast to [Figure 8](#), where we hold the gender constant and analyze across sanction statuses, in Figure S18 in [Multimedia Appendix 1](#), we hold the sanction status constant and analyze across genders. This shows the differences in reviews for both sanctioned women and men and unsanctioned women and men. For both specialties, we found small gender differences among unsanctioned physicians. However, among sanctioned physicians, the differences are large: women, especially female internists, are disproportionately reviewed in a more negative manner. Patients in the OBGYN department tend to review sanctioned male physicians more emotionally, in both positive and negative terms (with the exception of disgust and negativity, for which sanctioned female OBGYN physicians score high on average).

Overall, we conclude that emotional scoring analysis adds a layer of depth to our understanding of the differences among lexical reviews of physicians. The differences between the specialties are even more fascinating—although we are unable to discern major differences between the specialties regarding general word composition of the lexicons, the emotional discrepancies between internal medicine reviews (written by a mix of patients) and OBGYN reviews (written by mostly female patients) are extremely clear. Holding everything else constant, internal medicine reviews of male physicians tend to be largely more emotional, regardless of whether that emotion is positive or negative. Reviews of female OBGYN physicians tend to be much more negative.

When we add sanction status to the analysis, the dynamic becomes more complex. In internal medicine, there are no notable differences in the emotional scores of sanctioned men and unsanctioned men. In contrast, reviews of sanctioned female physicians in internal medicine show negative emotion more prominently than those of unsanctioned female physicians. The biggest difference between emotional scores in this entire analysis is between unsanctioned and sanctioned male OBGYN physicians—sanctioned male OBGYN physicians receive the most negative reviews in any subset of data analyzed. When comparing sanctioned women directly with sanctioned men, sanctioned female internal medicine physicians are reviewed much more negatively than sanctioned male internal medicine physicians, but reviews of sanctioned male OBGYN physicians are more emotional overall, regardless of whether the emotion is positive or negative.

## Discussion

### Overview

In this study, we analyzed web-based reviews of physicians and how they differ based on physicians' gender. We further sought to understand the complex interaction among the physician's web-based score (rating), whether they are sanctioned by a state medical board, and gender, as revealed in the content of the web-based reviews. To investigate this interaction, we implemented paragraph vector techniques to identify words that are specific to and indicative of the separate metadata cuts. Then, we enriched these findings by using the NRC

word-emotion association lexicon to assign emotional scores to 3 segments: gender, gender and sanction, and gender and rating.

### Principal Findings

Our findings shed light on the different criteria by which patients evaluate male and female physicians, and they highlight the disparity in severity with which patients review male and female physicians. When we analyze the ratings of male and female physicians while holding the rating range constant, it becomes clear that women are more likely to be evaluated on their interpersonal bedside manner, whereas men are more likely to be evaluated based on their perceived technical skills and performance. This pattern holds when analyzing reviews of low-rated or medium-rated male physicians—the lexical content of their reviews is still much more likely to convey high praise, whereas women are much more likely to be severely criticized. The dynamic is further exacerbated among men and women who are sanctioned. It is much more difficult to discern a review of a sanctioned man from the review of an unsanctioned man by the content of the written review alone, whereas for women, there is a stark contrast, and female physicians are evaluated much more harshly if they are sanctioned. The insight gained by analyzing sanctioned physicians is an important contribution of this study. There are baseline differences between how male and female physicians are perceived, but those differences are greatly magnified when the service quality is low. Sanctioned men still receive glowing reviews, whereas sanctioned women experience large reputational penalties when they deliver low-quality care or behave inappropriately.

It is essential to understand not only the quantitative differences in how and why female and male physicians are evaluated but also the qualitative aspect of those differences. Contributing to this qualitative understanding, our findings elucidate the gender-driven difference in bases for evaluations of physicians by patients. Most notably, we did not see differences in the emotional language used for sanctioned and unsanctioned male physicians, whereas female physicians who will be sanctioned have consistently more negative emotion associated with their reviews.

### Comparison With Previous Studies

An expanding stream of literature shows significant gender bias in ratings, perhaps most egregiously in a case in which changing the name of an anonymous teaching assistant from male to female lowered the average review score [29]. Our study contributes to the growing literature on how web-based medical reviews are biased by gender, highlighting that in web-based reviews, women are more likely to receive negative reviews, obtain low scores, and be judged on criteria not directly related to their skills as a physician (eg, diagnostic abilities) [20,21]. We make a unique contribution by examining how physicians who are sanctioned for inappropriate behavior, negligence, or malpractice are penalized for low-quality service.

### Limitations

Our results are subject to a few limitations imposed by the data. First, we only have review data and do not know the actual quality of care delivered (except care by sanctioned physicians,

which we know is more likely to be poor). We do not know the types of services received, and we do not know the patient outcomes. We tried to account for these unknowns by averaging all patient reviews for each physician and comparing physicians within subspecialties, which should control for much of the variation in services provided. However, if the medical services provided within a subspecialty systematically differ between genders, there may still be some residual confounding. Second, our data set does not contain physicians' race or ethnicity, which is another potential dimension of review bias. Future studies can investigate the possibility of racial and gender bias. Third, in our data, sanctioned physicians' reviews before the sanction date were combined; therefore, we could not explore the commonality or information signals provided by no-text reviews or by the length of individual reviews. Fourth, owing to the small number of sanctioned physicians, we represented the presence or absence of sanctions with a binary indicator; however, sanction severity varies. Therefore, future studies can focus on sanction severity to provide a more detailed and nuanced analysis of reviews. Finally, we acknowledge that the data are a decade old at the time of publication, meaning that

if there have been sociological changes in patients' views and behavior related to physician's gender, our results will not capture those recent developments.

## Conclusions

The role and influence of web-based reviews may grow as medicine becomes increasingly computerized, a shift that has only been accelerated by the COVID-19 pandemic. As telemedicine expands in scope and prevalence, proximity becomes less of a limiting factor in selecting a physician; therefore, patients will rely more on web-based reviews to guide their physician choices. Given this growing role of reviews in physician selection, action needs to be taken to ensure that they are fair and balanced. Although awareness is the first step, websites and apps that feature or contain physician reviews should also follow best practices for mitigating gender and racial bias in those reviews. For instance, as previous studies have shown, asking specific questions rather than providing open-ended boxes for reviews can reduce bias [30]. Similarly, highlighting the potential for unconscious bias [31] and providing a rubric for evaluations [32] can also help web-based platforms to mitigate biases in physician reviews.

## Authors' Contributions

The authors' order of contribution was as follows: JB, CC, MVB, and DA. JB drafted the *Methods* and *Results* sections and conducted the final data and modeling analysis. CC conducted the initial model fitting and exploratory analysis. MVB and DA collaborated in the overall guidance and direction of the paper, acquired the data, and edited the manuscript. DA drafted the *Introduction* and *Discussion* sections. All the authors approved the final manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Additional results, charts, and tables.

[[PDF File \(Adobe PDF File\), 737 KB - formative\\_v6i9e34902\\_app1.pdf](#)]

## References

1. Holliday AM, Kachalia A, Meyer GS, Sequist TD. Physician and patient views on public physician rating websites: a cross-sectional study. *J Gen Intern Med* 2017 Jun;32(6):626-631 [FREE Full text] [doi: [10.1007/s11606-017-3982-5](https://doi.org/10.1007/s11606-017-3982-5)] [Medline: [28150098](https://pubmed.ncbi.nlm.nih.gov/28150098/)]
2. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. *J Med Internet Res* 2013 Feb 01;15(2):e24 [FREE Full text] [doi: [10.2196/jmir.2360](https://doi.org/10.2196/jmir.2360)] [Medline: [23372115](https://pubmed.ncbi.nlm.nih.gov/23372115/)]
3. Xu Y, Armony M, Ghose A. The interplay between online reviews and physician demand: an empirical investigation. *Manag Sci* 2021 Dec;67(12):7344-7361. [doi: [10.1287/mnsc.2020.3879](https://doi.org/10.1287/mnsc.2020.3879)]
4. Hu N, Liu L, Sambamurthy V. Fraud detection in online consumer reviews. *Decis Support Syst* 2011 Feb;50(3):614-626. [doi: [10.1016/j.dss.2010.08.012](https://doi.org/10.1016/j.dss.2010.08.012)]
5. Lantzy S, Anderson D. Can consumers use online reviews to avoid unsuitable doctors? Evidence from rateMDs.com and the Federation of State Medical Boards. *Decis Sci* 2020 Aug;51(4):962-984. [doi: [10.1111/dec.12398](https://doi.org/10.1111/dec.12398)]
6. Lieber R. The Web Is Awash in Reviews, but Not for Doctors. Here's Why. *The New York Times*. 2012 Mar 9. URL: <https://www.nytimes.com/2012/03/10/your-money/why-the-web-lacks-authoritative-reviews-of-doctors.html> [accessed 2021-10-01]
7. McGrath RJ, Priestley JL, Zhou Y, Culligan PJ. The validity of online patient ratings of physicians: analysis of physician peer reviews and patient ratings. *Interact J Med Res* 2018 Apr 09;7(1):e8 [FREE Full text] [doi: [10.2196/ijmr.9350](https://doi.org/10.2196/ijmr.9350)] [Medline: [29631992](https://pubmed.ncbi.nlm.nih.gov/29631992/)]
8. Dunivin Z, Zadunayski L, Baskota U, Siek K, Mankoff J. Gender, soft skills, and patient experience in online physician reviews: a large-scale text analysis. *J Med Internet Res* 2020 Jul 30;22(7):e14455 [FREE Full text] [doi: [10.2196/14455](https://doi.org/10.2196/14455)] [Medline: [32729844](https://pubmed.ncbi.nlm.nih.gov/32729844/)]

9. Roter D, Lipkin Jr M, Korggaard A. Sex differences in patients' and physicians' communication during primary care medical visits. *Med Care* 1991 Nov;29(11):1083-1093. [doi: [10.1097/00005650-199111000-00002](https://doi.org/10.1097/00005650-199111000-00002)] [Medline: [1943269](https://pubmed.ncbi.nlm.nih.gov/1943269/)]
10. Jefferson L, Bloor K, Birks Y, Hewitt C, Bland M. Effect of physicians' gender on communication and consultation length: a systematic review and meta-analysis. *J Health Serv Res Policy* 2013 Oct;18(4):242-248 [FREE Full text] [doi: [10.1177/1355819613486465](https://doi.org/10.1177/1355819613486465)] [Medline: [23897990](https://pubmed.ncbi.nlm.nih.gov/23897990/)]
11. Greenwood BN, Carnahan S, Huang L. Patient-physician gender concordance and increased mortality among female heart attack patients. *Proc Natl Acad Sci U S A* 2018 Aug 21;115(34):8569-8574 [FREE Full text] [doi: [10.1073/pnas.1800097115](https://doi.org/10.1073/pnas.1800097115)] [Medline: [30082406](https://pubmed.ncbi.nlm.nih.gov/30082406/)]
12. Hao H, Zhang K. The voice of Chinese health consumers: a text mining approach to Web-based physician reviews. *J Med Internet Res* 2016 May 10;18(5):e108 [FREE Full text] [doi: [10.2196/jmir.4430](https://doi.org/10.2196/jmir.4430)] [Medline: [27165558](https://pubmed.ncbi.nlm.nih.gov/27165558/)]
13. López A, Detz A, Ratanawongsa N, Sarkar U. What patients say about their doctors online: a qualitative content analysis. *J Gen Intern Med* 2012 Jun;27(6):685-692 [FREE Full text] [doi: [10.1007/s11606-011-1958-4](https://doi.org/10.1007/s11606-011-1958-4)] [Medline: [22215270](https://pubmed.ncbi.nlm.nih.gov/22215270/)]
14. Thawani A, Paul MJ, Sarkar U, Wallace BC. Are online reviews of physicians biased against female providers? In: *Proceedings of the 4th Machine Learning for Healthcare Conference*. 2019 Presented at: PMLR '19; August 8-10, 2019; Ann Arbor, MI, USA p. 406-423.
15. Marrero K, King E, Fingeret AL. Impact of surgeon gender on online physician reviews. *J Surg Res* 2020 Jan;245:510-515. [doi: [10.1016/j.jss.2019.07.047](https://doi.org/10.1016/j.jss.2019.07.047)] [Medline: [31446193](https://pubmed.ncbi.nlm.nih.gov/31446193/)]
16. Kordzadeh N. Investigating bias in the online physician reviews published on healthcare organizations' websites. *Decis Support Syst* 2019 Mar;118:70-82. [doi: [10.1016/j.dss.2018.12.007](https://doi.org/10.1016/j.dss.2018.12.007)]
17. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. *J Am Med Inform Assoc* 2014;21(6):1098-1103 [FREE Full text] [doi: [10.1136/amiajnl-2014-002711](https://doi.org/10.1136/amiajnl-2014-002711)] [Medline: [24918109](https://pubmed.ncbi.nlm.nih.gov/24918109/)]
18. Rivas R, Montazeri N, Le NX, Hristidis V. Automatic classification of online doctor reviews: evaluation of text classifier algorithms. *J Med Internet Res* 2018 Nov 12;20(11):e11141 [FREE Full text] [doi: [10.2196/11141](https://doi.org/10.2196/11141)] [Medline: [30425030](https://pubmed.ncbi.nlm.nih.gov/30425030/)]
19. Wartena C, Sander U, Patzelt C. Sentiment independent topic detection in rated hospital reviews. In: *Proceedings of the 13th International Conference on Computational Semantics - Short Papers*. 2019 Presented at: IWCS '19; May 23-27, 2019; Gothenburg, Sweden p. 59-64. [doi: [10.18653/v1/w19-0509](https://doi.org/10.18653/v1/w19-0509)]
20. Li S, Lee-Won RJ, McKnight J. Effects of online physician reviews and physician gender on perceptions of physician skills and Primary Care Physician (PCP) selection. *Health Commun* 2019 Oct;34(11):1250-1258. [doi: [10.1080/10410236.2018.1475192](https://doi.org/10.1080/10410236.2018.1475192)] [Medline: [29792519](https://pubmed.ncbi.nlm.nih.gov/29792519/)]
21. Physician Data Center Query. Federation of State Medical Boards. 2018. URL: <https://www.fsmb.org/PDC/> [accessed 2022-08-26]
22. Manning CD, Schütze H. *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press; 1999.
23. Lewis DD, Yang Y, Rose TG, Li F. Rcv1: a new benchmark collection for text categorization research. *J Mach Learn Res* 2004;5(Apr):361-397.
24. Rijsbergen CJ, Robertson SE, Porter MF. *New Models in Probabilistic Information Retrieval*. Vol. 5587. London, UK: British Library Research and Development Department; 1980.
25. Porter MF. An algorithm for suffix stripping. *Program* 1980;14(3):130-137. [doi: [10.1108/eb046814](https://doi.org/10.1108/eb046814)]
26. Le Q, Mikolov T. Distributed representations of sentences and documents. In: *Proceedings of the 31st International Conference on Machine Learning*. 2014 Presented at: PMLR '14; June 21-26, 2014; Beijing, China p. 1188-1196.
27. Mohammad SM, Turney PD. Crowdsourcing a word-emotion association lexicon. *Comput Intell* 2013 Aug;29(3):436-465. [doi: [10.1111/j.1467-8640.2012.00460.x](https://doi.org/10.1111/j.1467-8640.2012.00460.x)]
28. Lau JH, Baldwin T. An empirical evaluation of doc2vec with practical insights into document embedding generation. arXiv Preprint posted online July 19, 2016 [FREE Full text] [doi: [10.48550/ARXIV.1607.05368](https://doi.org/10.48550/ARXIV.1607.05368)]
29. MacNell L, Driscoll A, Hunt AN. What's in a name: exposing gender bias in student ratings of teaching. *Innov High Educ* 2014 Dec 5;40(4):291-303. [doi: [10.1007/s10755-014-9313-4](https://doi.org/10.1007/s10755-014-9313-4)]
30. Castilla EJ. Gender, race, and meritocracy in organizational careers. *AJS* 2008 May;113(6):1479-1526. [doi: [10.1086/588738](https://doi.org/10.1086/588738)] [Medline: [19044141](https://pubmed.ncbi.nlm.nih.gov/19044141/)]
31. Peterson DA, Biederman LA, Andersen D, Ditonto TM, Roe K. Mitigating gender bias in student evaluations of teaching. *PLoS One* 2019 May 15;14(5):e0216241 [FREE Full text] [doi: [10.1371/journal.pone.0216241](https://doi.org/10.1371/journal.pone.0216241)] [Medline: [31091292](https://pubmed.ncbi.nlm.nih.gov/31091292/)]
32. Uhlmann EL, Cohen GL. Constructed criteria: redefining merit to justify discrimination. *Psychol Sci* 2005 Jun;16(6):474-480. [doi: [10.1111/j.0956-7976.2005.01559.x](https://doi.org/10.1111/j.0956-7976.2005.01559.x)] [Medline: [15943674](https://pubmed.ncbi.nlm.nih.gov/15943674/)]

## Abbreviations

- LDA:** latent Dirichlet allocation  
**NRC:** National Research Council Canada  
**OBGYN:** obstetrics and gynecology

*Edited by A Mavragani; submitted 17.11.21; peer-reviewed by R McGrath, J Fan, H Stevens; comments to author 24.01.22; revised version received 26.04.22; accepted 17.07.22; published 08.09.22.*

*Please cite as:*

*Barnett J, Bjarnadóttir MV, Anderson D, Chen C*

*Understanding Gender Biases and Differences in Web-Based Reviews of Sanctioned Physicians Through a Machine Learning Approach: Mixed Methods Study*

*JMIR Form Res 2022;6(9):e34902*

*URL: <https://formative.jmir.org/2022/9/e34902>*

*doi: [10.2196/34902](https://doi.org/10.2196/34902)*

*PMID: [36074543](https://pubmed.ncbi.nlm.nih.gov/36074543/)*

©Julia Barnett, Margrét Vilborg Bjarnadóttir, David Anderson, Chong Chen. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 08.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# The Helpfulness of Web-Based Mental Health and Well-being Forums for Providing Peer Support for Young People: Cross-sectional Exploration

Emily Banwell<sup>1</sup>, BSc, MSc; Terry Hanley<sup>1</sup>, BSc, MA, MSc, PhD; Santiago De Ossorno Garcia<sup>2</sup>, BSc, MRes, MSc, PhD; Charlotte Mindel<sup>2</sup>, BSc, MSc; Thomas Kayll<sup>2</sup>, BSc; Aaron Sefi<sup>2,3</sup>, BA, MA

<sup>1</sup>Manchester Institute of Education, University of Manchester, Manchester, United Kingdom

<sup>2</sup>Kooth Plc, London, United Kingdom

<sup>3</sup>Department of Psychology, University of Exeter, Exeter, United Kingdom

**Corresponding Author:**

Emily Banwell, BSc, MSc

Manchester Institute of Education

University of Manchester

Ellen Wilkinson Building

Oxford Road

Manchester, M15 6JA

United Kingdom

Email: [emily.banwell@manchester.ac.uk](mailto:emily.banwell@manchester.ac.uk)

## Abstract

**Background:** Young people are increasingly seeking out web-based support for their mental health and well-being. Peer support forums are popular with this age group, with young individuals valuing the fact that the forums are available 24/7, providing a safe and anonymous space for exploration. Currently, little systematic evaluation of the helpfulness of such forums in providing support has been conducted.

**Objective:** This study examined the helpfulness of the support offered within web-based mental health and well-being peer support forums for young people. It specifically investigated the self-reported user ratings of helpfulness reported through the completion of a developing experience measure. The ratings will be used to consider further development of the measure and reflect upon the overall helpfulness of the forums as indicated by the reported scores.

**Methods:** The study used routinely collected practice-based outcome data from web-based mental health forums for young people. These forums are hosted by the UK-based web-based therapy and support service, Kooth. A cross-sectional design was used to explore—using a range of inferential statistical measures—the outcomes reported by those accessing the forums using a Peer Online Community Experience Measure (POCEM). To consider the helpfulness in general, 23,443 POCEMs completed in 2020 were used. A second data set of 17,137 completed POCEMs from the same year was used to consider whether various engagement indicators had an impact upon the helpfulness rating.

**Results:** Female users aged between 11 and 16 years predominantly completed the POCEM. This is in keeping with the majority of those using the service. In total, 74.6% (8240/11,045) of the scores on the POCEM indicated that the individuals found the posts helpful. An ANOVA indicated that male users were more likely to report obtaining intrapersonal support, whereas female users obtained interpersonal support. Furthermore, the POCEM scores reflected the internal consistency of the measure and provided an insight into the way that young people made use of the peer support resource; for instance, posts that were rated more helpful were correlated with spending longer time reading them, and the topics discussed varied throughout the day with more mental health issues being discussed later at night.

**Conclusions:** The results seem to demonstrate that, overall, the young people involved in this study found web-based peer support helpful. They indicate that peer support can provide an important strand of care within a supportive mental health ecosystem, particularly during time periods when in-person support is typically closed. However, limitations were noted, suggesting that caution is needed when interpreting the results of this study. Although such services are incredibly well used, they have received little research attention to date. As such, further investigation into what constitutes helpful and unhelpful peer support is needed.

**KEYWORDS**

adolescent mental health; peer support; web-based forums; web-based mental health

## Introduction

### Background

Several definitions of the term *peer support* have been proposed; yet, these can be unified in the sense that “peer support occurs between people who share similar life experiences and as a result can provide each other with reciprocal support...which professionals and/or others who have not endured the same difficult situations may not be able to” [1]. Digital web-based communities are an illustration of how peer support can be facilitated. Such communities are particularly pertinent for children and young people (CYP), who are turning to the internet with increasing frequency when seeking psychological and emotional support [2]. In 2017, 99% of the CYP aged 12 to 15 years spent approximately 21 hours a week on the web, with this figure seeming to steadily increase year on year [3,4]. Therefore, it is easy to comprehend how the practicality of web-based communities such as discussion forums is attractive to CYP [5]. This became especially true at the onset of the COVID-19 pandemic in 2020 when digital mental health support demand for CYP sharply increased [6,7]. Unlike in-person support, web-based forums are available 24/7 [8] and provide safe, comfortable platforms that are geographically unrestricted [5]. The 24/7 element of peer support forums may be especially useful for adolescents because young people’s circadian rhythms shift toward a later bedtime, with wakeful activity occurring late into the night [9,10]. Interestingly, those who experience such a shift are more likely to experience emotional and behavioral issues [11].

Even when forums are professionally led and moderated, web-based peer support is not intended to replace services provided by qualified therapists [12]. However, the fact that it is frequently available as an adjunct to other forms of mental health support [5,13] means that one can obtain additional, lower-level support that can reduce feelings of isolation [12] and normalize experiences [14]. Web-based peer support overcomes the physical obstacles that can often prevent CYP from building social relationships with others who share their experiences, allowing them to connect with peers whom they would not otherwise meet in *real life* [15]. Mental health discussion forums are particularly valued by young people who perceive themselves as lacking social skills [16]. An additional interesting finding within this body of research was that despite concerns about perpetuation of dangerous behaviors such as self-harm when young people experiencing a mental health crisis are united in a forum space, the discussion remained predominantly focused on safety and avoidance strategies [17]. This fits in with observations that people prefer sharing clinical, rather than personal and potentially identifiable information, in web-based peer support settings [18].

There are a number of challenges associated with the use of web-based forums for CYP to discuss mental health, such as ensuring user safety, building feelings of trust [5], and reducing

the possibility of aggressive or unfriendly behavior that can occur as a result of user anonymity [19]. Despite these considerations, the benefits uncovered thus far in the literature mean that more in-depth examination into their use is warranted. An outline of the literature focusing on patterns of forum engagement is provided in the next paragraph.

Within forums, support can be either directive or nondirective in nature. A directive approach comprises receipt or provision of explicit advice on how to overcome an issue, and a nondirective approach involves supportively sharing experiences [14]. CYP appreciate the anonymity and privacy of peer support forums in that they find talking easier and feel less judged as a consequence [5]. This level of anonymity could play a key role in how web-based communities are built and the subsequent level of benefit that can be gained from interacting in this format. So far, findings examining this idea have been mixed. Posts on the web-based *question-and-answer* site Quora were rated as equally useful whether submitted anonymously or otherwise, and interestingly, politeness of content did not vary with anonymity [20]. Contrastingly, other findings suggested that registered forum users tended to post higher-quality posts than those who posted anonymously [21]. Ensuring that content is useful is important if web-based communities or forums are to provide a valuable and engaging method of peer support. However, to do this we must identify the factors that allow us to call a peer support forum post *helpful*. One study of interest identified several factors that contribute to forum post performance (used here as a proxy for quality) of web-based forum threaded discussions [21]. These ranged from surface-level qualities such as thread length and reply latency to less quantifiable qualities such as the perceived authority of the poster. The study suggested that factors such as thread length, or number of comments, as indicators of participative engagement can be effective determinants of thread quality; for example, they can indicate how effectively a thread answered a question or explored a topic. The number of unique users and repeat users posting on a thread are proxies for wide-reaching engagement and a desire to continue discussing a topic, respectively [21]. To our knowledge, the study by Lee et al [21] is the only one that examined forum value at such a granular level. However, as a news and current affairs forum was the platform of focus, the unique nature of peer support forums means that transferability of the study’s findings to CYP mental health forums cannot be completely assumed. There have been mixed findings regarding the question of whether factors such as number of log-ins, view counts, or page clicks, in spite of their suggested links with participation, can actually be considered true proxies of engagement. Log-in frequency was linked to engagement with web-based student learning in 1 study [22], whereas 2 similar studies found no such link [23,24]. In addition, one of these studies [24] found no link between engagement and session duration. However, all 3 studies focused on web-based learning. This again means that the findings

cannot automatically be applied to digital mental health web-based forums.

It is clear from the outlined research that there are numerous positive outcomes of web-based mental health forums and communities for CYP [5,12,14]. Despite this, no empirical evidence exists that relates directly to the helpfulness of forum content that is situated within a web-based CYP peer community for mental health and well-being. As a result, this study built on recent work (Mindel, C, unpublished data, November 2021) where a Peer Online Community Experience Measure (POCEM) was developed in conjunction with a UK-based digital therapy platform for CYP. The recent preprint of the study by Mindel et al (Mindel, C, unpublished data, November 2021) detailed the POCEM's development and piloting, as well as the beginning of the investigation of how users interacted with the measure. However, further evaluation of the POCEM within nonpilot routine service use is needed, not only in terms of how the measure itself is engaged with but also what the measure can tell us about patterns of interaction with web-based community peer support at a wider level.

### Research Questions

This study aimed to explore user-rated measurement of the helpfulness of web-based peer support forum content by investigating the following research questions:

1. What is represented in the outcomes of an experience measure (POCEM) that claims to rate the helpfulness of web-based community content?
2. What trends and commonalities exist within posts that are deemed either helpful or less helpful by CYP?

## Methods

### Design

This study used an inferential cross-sectional design [25] to identify and explore the factors that underpin the outcomes of the POCEM. Scores from routinely collected completed POCEMs were compared with several other user-level variables to explore the patterns of use stipulated by the research questions.

### Setting

The data for this study came from Kooth, which is a UK-based digital therapy and support service for CYP aged 10 to 25 years. The service provides a predominantly humanistic and idiographic approach to counseling, where users are actively involved in every step of their therapeutic journey [26]. A variety of structured or self-directed help is available through the service, with peer support community forums providing just 1 form of this available help. Service users can create and interact with forum posts on a wide variety of topics, ranging from mental health and hobbies and interests to sexuality and relationships. Worker-initiated activities and discussion threads also form part of this community to initiate discussions and encourage engagement.

No referral or joining a waiting list is needed to use Kooth, which is a free-of-charge service. However, service access is, at present, contract limited. This means that access is restricted

to certain geographical regions as governed by whole-population contracts with local National Health Service trusts. Kooth can also be commissioned by other organizations, such as local authorities, for use with certain subsets of the population for whom a need for provision is recognized. Although the web-based community forums are accessible 24/7, live counselors and moderators are only active until 10 PM.

### Ethical Considerations

All data used in this study were collected from service users who had consented to their data being used for research purposes. As this was routinely collected information, the research process did not influence the data [27]. In addition, because the resulting data were anonymous and devoid of personally identifiable information, ethical review from the lead author's institution was not required [28]. Even so, ethical principles and good practice guidelines for managing information of this nature were followed [29].

### Measures and Data Sets

#### Overview of the POCEM

The POCEM is a 3-step web-based experience measure designed for users to rate content within the Kooth community. It aims to measure satisfaction with, and quality of, web-based community resources in relation to how well they met the expectations of the service user completing the measure. Any time a service user engages with a community item such as a forum post or an activity, they are, as a first step, asked how helpful they found it by rating it on a Likert scale (-2: No; -1: Not really; 0: Not sure; 1: A bit; 2: Loads). The second and third steps are explained in the Data Set 1 section.

#### Data Set 1

Two data sets were used to investigate the research aims. For the first aim—to examine the user-rated measurement of helpfulness and experiences gained from community forum interaction—a cross-sectional extraction of every POCEM completed in 2020 was performed. This data set of 23,443 POCEMs was completed by 11,045 unique users. Each user completed a mean of 2.12 (SD 2.64) POCEMs. This data extraction captured the second and third steps of the 3-step measure. In the second step, after assigning a numerical score to a post, the service user was prompted to indicate a reason for their rating, choosing from 4 options to reflect the nature of their experience. This was optional, and their Likert rating choice was recorded even if they chose not to engage further. In addition, if the user selected *not sure*, the interaction with the measure would end at this point. The 4 experience domain options to choose from are based on 4 high-level outcomes [30] that were devised from the different types of support that are commonly sought (refer to the *High-level support represented by each domain* column in Table 1). Third and last, users were then given the option to provide more information about their experience by selecting from a range of statements that correspond to the domain that they had selected. These additional statements (refer to the *Additional selectable statements* column in Table 1) represent indicators of outcomes of a positive community experience, and users could choose one or more of these. This data set also captured the



demographics of age, gender, and ethnicity of users completing the measure, as well as the topic category to which the POCEM-rated content related; for example, *mental health*, *bullying*, or *friends*.

**Table 1.** The selectable domains and their corresponding statements that are available in the Peer Online Community Experience Measure and the high-level support outcomes that they represent.

Service user–experience domains	Additional selectable statements under each domain	High-level support represented by each domain
“I want information about something important to me” (Important to me)	<ul style="list-style-type: none"> <li>• I got information that helped me learn about myself</li> <li>• The information I received today was helpful to my problem</li> <li>• I now know what I need to do to feel better</li> <li>• I learned something new today</li> <li>• I now know that others have the same experiences as me</li> </ul>	Informational intrapersonal support
“I want to learn some skills to try with other people” (Learn skills)	<ul style="list-style-type: none"> <li>• I have learned how to express myself</li> <li>• I have learned enough to make a positive change</li> <li>• I have developed skills to open up more</li> <li>• I now have knowledge and skills to help others</li> <li>• I have learned how to support others</li> </ul>	Informational interpersonal support
“I want to explore more about how I relate to other people” (Relate to others)	<ul style="list-style-type: none"> <li>• I feel safe in the Kooth community</li> <li>• I felt connected to someone</li> <li>• I feel that I’m just as valuable as others</li> <li>• I know who to ask for help</li> <li>• I feel motivated to give advice to others</li> <li>• It feels good not to be judged</li> </ul>	Emotional interpersonal support
“I want to understand myself more” (Understand myself)	<ul style="list-style-type: none"> <li>• I felt accepted</li> <li>• I now feel more hopeful</li> <li>• I now feel able to ask for support outside of Kooth</li> <li>• My problems now feel more manageable</li> <li>• I am now able to find solutions to my problems</li> <li>• I now want to make changes in my life</li> <li>• I no longer feel alone</li> </ul>	Emotional intrapersonal support

## Data Set 2

The second data set, a subset of data set 1, comprised 17,137 POCEMs completed by 10,612 unique users. In this data set, an average of 1.62 (SD 1.89) POCEMs were completed per user. It was used to explore the nuances and commonalities that exist among posts that were rated a certain way, as per the second research aim. This data set was a cross-sectional data extraction of those POCEMs completed within 1 month of their corresponding community forum post being submitted. As a clarifying example, a POCEM that was completed on April 21 that related to a forum post submitted on March 22 would be included in this extraction; however, a POCEM completed on April 23 would not be included. This limit was set because engagement with posts tends to become sparse when they are >1 month old. As in the first data set, a measurement score as well as the category topic of the community forum post were captured. The number of comments on, and views of, a post at the time of each POCEM completion were ascertained too. Owing to the interface of Kooth community forums, these features related to what a user can see when they enter the forum environment. The amount of time each user spent interacting with the post (in minutes) before they submitted the measurement scores, the average score of already submitted POCEMs, and the age of each post (in days) when each POCEM was completed were also recorded. In the Kooth web-based community, forum posts do not move to the top of the page

when a new comment is submitted. If the age of posts is linked to POCEM completion likelihood and score, it could provide useful insight into how users search for, and engage with, posts.

## Data Analysis

### Data Set 1

Frequency tables were produced to explore the demographics of those who completed POCEMs, the distribution of scores, and the topics to which the content related. The split of gender and ages across the 4 experience domains (Table 1) were explored with 1-way Welch ANOVAs, which were chosen because the data violated the *homogeneity of variance* assumption [31]. For these tests, age was divided into  $\leq 13$  years and  $\geq 14$  years. This boundary was chosen owing to the developmental transition into adolescence that is commonly found to occur around this age and because of the finding that the peak age of onset of any diagnosable mental health issue is 14.5 years [32]. Consequently, these 2 age groups may approach peer support in different ways. We chose not to further split this exploration by topic category because these varied so drastically in representation; for example, 26.49% (6210/23,443) of the POCEMs that were completed related to the mental health topic, but only 0.04% (10/23,443) of the POCEMs that were completed related to the independence topic.”

Heat maps were produced to explore which pairs of corresponding statements (refer to the *Additional selectable*

statements column in Table 1) were the most frequently selected together for each domain. Heat maps are a form of data visualization where tiles in a matrix are shaded based on the combined value of the corresponding axes. They are often shaded from red (low frequency or value) to green (high frequency or value) [33]. They were incorporated into this study to explore any potential patterns of consistencies or contrasts that exist in the service user experience. This was carried out by using Microsoft Excel's COUNTIF function to create dummy variables that detected the presence of each statement across each individual POCEM, then using the crosstab function of SPSS software (version 25.0; IBM Corp) to obtain counts of pairs to be inputted into the heat maps.

### Data Set 2

A simple correlation matrix explored the potential linear relationships among the various qualities of a community post at the time of POCEM completion. Next, a new variable was computed to split the data based on the time of day that each POCEM was completed to investigate how CYP use their free time outside of education or work to take care of their mental health and what patterns of completion followed after 10 PM when live counselors are offline. Using this variable, 1-way Welch ANOVAs investigated whether POCEM scores, time spent reading a post within the web-based community forum, and the post's age at POCEM completion varied depending on the time of day. Frequency tables examined trends relating to the topics interacted with at different times of day, with further ANOVAs used to explore points of interest by isolating topic categories through the computation of dummy variables.

## Results

In the following sections, we detail the findings of the investigations outlined in the Methods section. Please note that where dummy variables were inputted into significance tests, means are not provided because they would represent transformations rather than meaningful reflections of the original data.

### Data Set 1

In total, 75.3% (17,653/23,443) of the POCEMs in this data set were completed by female users and 18.9% (4431/23,443) by male users, whereas 3.7% (867/23,443) and 2.1% (492/23,443) were completed by respondents who identified as gender fluid

and agender, respectively. In total, 91.2% (21,380/23,443) of the measures were completed by users aged between 11 and 16 years, which corresponds with secondary school age in the United Kingdom. Of the 11,045 users, 4628 (41.9%) were aged 12 or 13 years, and only 376 (3.4%) were aged >18 years. In terms of ethnicity, 79.79% (8813/11,045) of the sample identified as White (White British, White Irish, or "other White background"), which falls below the 86% reported in the 2011 England and Wales census [34].

In terms of helpfulness scores, 74.6% (8240/11,045) of the users selected 1 or 2, indicating that they found the community forum post helpful, whereas 13.5% (1491/11,045) were unsure, and 12% (1325/11,045) found the post unhelpful. POCEMs were completed on posts relating to 16 topics. In total, 26.49% (6210/23,443) of the POCEMs were completed for mental health-related posts, with the next most common being sex and relationships at 13.63% (3196/23,443).

Significant gender differences existed in the likelihood of choosing the domains *important to me* ( $F_{3,1037.5}=2.845$ ;  $P=.04$ ) and *understand myself* ( $F_{3,1041.54}=2.834$ ;  $P=.04$ ), with more male users than female users choosing these as reflective of their community experience. A difference was also found for the domain *relate to others* ( $F_{3,1040.28}=4.965$ ;  $P=.002$ ), but conversely, more female users than male users chose this. No gender differences existed in selecting the *learn skills* domain, and those identifying as gender nonconforming did not differ significantly from male users or female users in any of their choices. CYP aged  $\geq 14$  years were significantly more likely than those aged  $\leq 13$  years to choose *relate to others* ( $F_{1,13555.74}=82.66$ ;  $P<.001$ ). Those aged  $\leq 13$  years were more likely than those aged  $\geq 14$  years to select *important to me* ( $F_{1,13999.15}=12.05$ ;  $P=.001$ ), *learn skills* ( $F_{1,14207.18}=34.78$ ;  $P<.001$ ), and *understand myself* ( $F_{1,13741.59}=15.85$ ;  $P<.001$ ).

Figure 1 shows heat maps that show the concentration of the pairs of additional selectable statements that were chosen together under each service user-experience domain. The pairs that were selected together the most frequently are in darker green, with the red squares representing pairs chosen together less often. The heat maps contain data only from those POCEMs where at least two statements were selected (3241/23,443, 13.83%).

**Figure 1.** Heat maps showing the concentration of pairs of selectable statements. Those selected together with the highest frequency are dark green, with red indicating the lowest frequency. N/A: not applicable.

**Important to me**

	I learned something new today	I got information that helped me learn about myself	I know what I need to do to feel better	I now know others have the same experiences as me	The information I received today was helpful to my problem
I learned something new today	N/A				
I got information that helped me learn about myself	72	N/A			
I know what I need to do to feel better	75	81	N/A		
I now know others have the same experiences as me	72	94	89	N/A	
The information I received today was helpful to my problem	76	85	88	92	N/A

**Learn skills**

	I have learned how to express myself	I have developed skills to open up more	I have learned how to support others	I have learned enough to make a positive change	I now have knowledge and skills to help others
I have learned how to express myself	N/A				
I have developed skills to open up more	113	N/A			
I have learned how to support others	97	118	N/A		
I have learned enough to make a positive change	94	103	152	N/A	
I now have knowledge and skills to help others	102	108	176	158	N/A

**Relate to others**

	I feel safe in the Kooth community	I know who to ask for help	It feels good not to be judged	I felt connected to someone	I feel that I'm just as valuable as others	I feel motivated to give advice to others
I feel safe in the Kooth community	N/A					
I know who to ask for help	534	N/A				
It feels good not to be judged	1109	529	N/A			
I felt connected to someone	955	460	1037	N/A		
I feel that I'm just as valuable as others	639	350	715	659	N/A	
I feel motivated to give advice to others	656	221	702	688	411	N/A

**Understand myself**

	I felt accepted	I now feel more hopeful	I now feel able to ask for support outside of Kooth	My problems now feel more manageable	I am now able to find solutions to my problems	I now want to make changes in my life	I no longer feel alone
I felt accepted	N/A						
I now feel more hopeful	222	N/A					
I now feel able to ask for support outside of Kooth	51	54	N/A				
My problems now feel more manageable	161	186	50	N/A			
I am now able to find solutions to my problems	135	147	43	155	N/A		
I now want to make changes in my life	128	134	38	123	108	N/A	
I no longer feel alone	114	112	0	94	77	58	N/A

**Data Set 2**

Negligible yet significant correlations were found between the POCEM score and age of the post at time of measure completion ( $r_{17,135}=0.03$ ;  $P<.001$ ) and between the POCEM score and time spent on the post ( $r_{17,135}=0.08$ ;  $P<.001$ ).

Significant correlations were also found between the age of the post and the following variables: time spent on the post ( $r_{17,135}=0.16$ ;  $P<.001$ ), average score of previously submitted POCEMs on the post ( $r_{17,135}=0.42$ ;  $P<.001$ ), and total number of submitted measures on the post ( $r_{17,135}=0.40$ ;  $P<.001$ ).

The amount of time in minutes a user spent reading or otherwise interacting with a post before completing a POCEM was significantly correlated with the number of views that the post had at the time of POCEM completion ( $r_{17,135}=0.31$ ;  $P<.001$ ), number of comments at the time of POCEM completion ( $r_{17,135}=0.49$ ;  $P<.001$ ), average score of previously submitted POCEMs ( $r_{17,135}=0.28$ ;  $P<.001$ ), and total number of submitted POCEMs on the post ( $r_{17,135}=0.37$ ;  $P<.001$ ).

When the POCEMs were broken down per the time of day in which they were completed, 48.9% (8380/17,137) were

completed between 4 PM and 10 PM (*evening and night*), with 39.7% (6803/17,137) completed between 6 AM and 4 PM (*morning and afternoon*). In total, 11.4% (1954/17,137) of the POCEMs in the sample were completed between 10 PM and 6 AM (*late night*).

The POCEM scores did not vary depending on time of day ( $F_{4,7248.25}=0.841$ ;  $P=.499$ ). However, significant main effects of post age at time of POCEM completion ( $F_{4,7166.89}=31.14$ ;  $P<.001$ ) and time spent engaging with a post before POCEM completion ( $F_{4,7445.44}=3.52$ ;  $P=.007$ ) were found. Games-Howell post hoc tests revealed several significant differences (Tables 2 and 3).

When examining the topics of forum posts for which POCEMs were completed at different times of day, the most popular topics were checked for apparent patterns. In the *morning*, between 6 AM and noon, 22.9% (3924/17,137) of the POCEMs were related to a mental health-themed forum post. This percentage increased as the day progressed, reaching 33.6% (5758/17,137) at *late night* (Figure 2). However, hobbies or interests followed an opposite pattern. In the *morning*, 13% (2228/17,137) of the POCEMs were related to this topic, but by *late night*, this fell to 7.4% (1268/17,137; Figure 3).

Exploring these phenomena, measure completion for mental health-related posts significantly varied depending on time of day ( $F_{4,7167.85}=21.95$ ;  $P<.001$ ). Games-Howell post hoc tests showed that all pairwise comparisons were significant, besides night and late night ( $P=.22$ ) and afternoon and evening ( $P=.87$ ).

POCEM completion for hobbies- or interests-related posts also significantly varied depending on time of day ( $F_{4,7292.65}=16.26$ ;  $P<.001$ ). All pairwise comparisons were significant, besides night and late night ( $P=.99$ ), afternoon and evening ( $P=.11$ ), evening and night ( $P=.16$ ), and evening and late night ( $P=.09$ ).

**Table 2.** Games-Howell pairwise comparisons among times of day in terms of forum post age (in days) at the time of Peer Online Community Experience Measure completion.

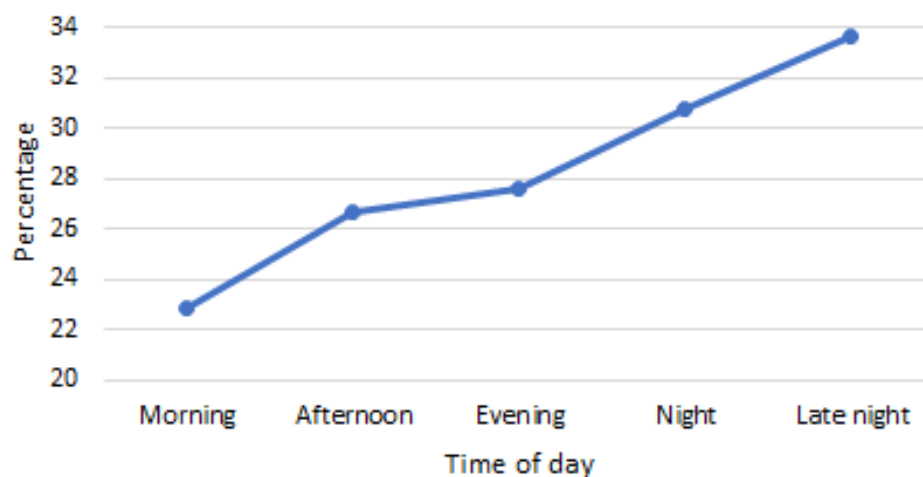
Time of day (mean post age in days) and pairwise comparison	<i>P</i> value
<b>Late night (mean 4.03, SD 4.35)</b>	
Morning	.003 <sup>a</sup>
Afternoon	.01 <sup>a</sup>
Evening	.12
Night	.003 <sup>a</sup>
<b>Morning (mean 4.48, SD 4.21)</b>	
Late night	.003 <sup>a</sup>
Afternoon	.99
Evening	<.001 <sup>a</sup>
Night	<.001 <sup>a</sup>
<b>Afternoon (mean 4.43, SD 4.40)</b>	
Late night	.01 <sup>a</sup>
Morning	.99
Evening	<.001 <sup>a</sup>
Night	<.001 <sup>a</sup>
<b>Evening (mean 3.76, SD 4.07)</b>	
Late night	.12
Morning	<.001 <sup>a</sup>
Afternoon	<.001 <sup>a</sup>
Night	.32
<b>Night (mean 3.58, SD 4.04)</b>	
Late night	.003 <sup>a</sup>
Morning	<.001 <sup>a</sup>
Afternoon	<.001 <sup>a</sup>
Evening	.32

<sup>a</sup>Values that met the significance threshold ( $P<.05$ ).

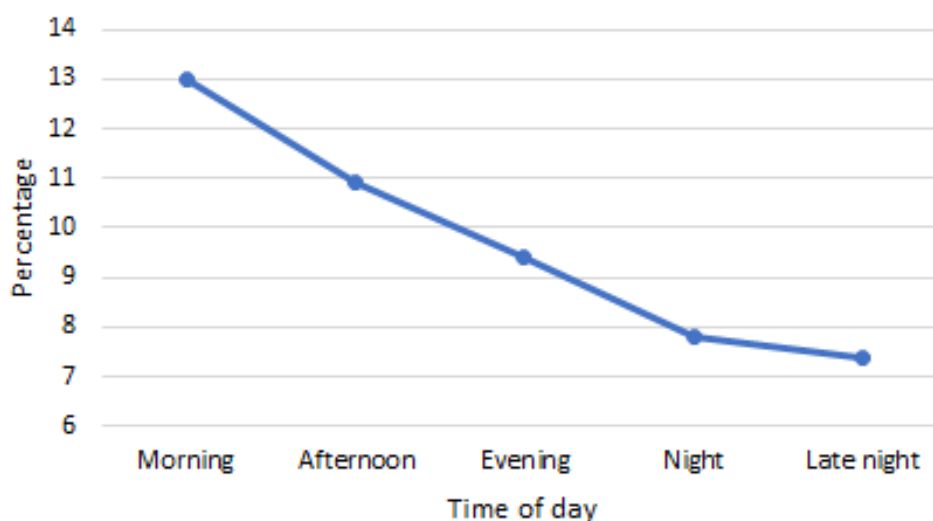
**Table 3.** Games-Howell pairwise comparisons among times of day in terms of time spent engaging with a post (in minutes) before Peer Online Community Experience Measure completion.

Time of day (mean time spent engaging with post in minutes) and pairwise comparison	<i>P</i> value
<b>Late night (mean 0.82, SD 0.62)</b>	
Morning	.02 <sup>a</sup>
Afternoon	.34
Evening	.03 <sup>a</sup>
Night	.89
<b>Morning (mean 0.88, SD 0.64)</b>	
Late night	.02 <sup>a</sup>
Afternoon	.62
Evening	.99
Night	.14
<b>Afternoon (mean 0.86, SD 0.67)</b>	
Late night	.34
Morning	.62
Evening	.78
Night	.87
<b>Evening (mean 0.88, SD 0.83)</b>	
Late night	.03 <sup>a</sup>
Morning	.99
Afternoon	.78
Night	.23
<b>Night (mean 0.84, SD 0.66)</b>	
Late night	.89
Morning	.14
Afternoon	.87
Evening	.23

<sup>a</sup>Values that met the significance threshold ( $P < .05$ ).

**Figure 2.** Percentage of total Peer Online Community Experience Measures completed that related to a mental health-themed post at different times of day.

**Figure 3.** Percentage of total Peer Online Community Experience Measures completed that related to a hobbies- or interests-themed post at different times of day.



## Discussion

### Overview

Web-based peer support, especially that pertaining to CYP mental health, has received very little direct research attention thus far despite its extensive use by this demographic. To close this gap, this study builds on a recent measure development paper (Mindel, C, unpublished data, November 2021) by providing novel insight into how CYP interact with a web-based community measure designed to rate the helpfulness of peer support forum content. From the information provided in the POCEMs, we were able to investigate the commonalities underpinning forum posts and web-based community content that were rated in certain ways.

### Outcomes of the POCEM

#### Demographics

In the first data set, user-rated measurements of helpfulness were examined. In total, 41.9% (9823/23,443) of the POCEMs within this extraction were completed by users aged 12 or 13 years. These ages are characterized by the transition from childhood into adolescence. Many social, educational, and biological changes affect a young person's life at this time [35]. In addition, approximately half of mental health conditions are commonly believed to begin before the age of 14 years [36], and the peak age of onset of any diagnosable mental health difficulty is 14.5 years [32]. This age clearly represents a crucial period for which attention to supportive prevention should be paramount. Ensuring that services, including those that are web based, provide targeted support for the specific issues faced by this age group may help with this. It is also worth noting that the low percentage of users aged >18 years in the sample (376/11,045, 3.4%) could be explained by this age group's eligibility for Kooth's adult-specific digital therapy platform, Qwell.

We also found that considerably more POCEMs were completed by female users (17,653/23,443, 75.3%) than male users (4431/23,443, 18.9%). Data from the Millennium Cohort Study

found that emotional mental health symptoms increased from 12% to 18% for girls aged between 11 and 14 years, although prevalence for boys stayed the same [35]. Although this higher susceptibility might have partial responsibility for the gender difference in use that we found, the magnitude of the difference could suggest that other factors are also in play. Gender differences in help-seeking behavior are well documented, with women more likely to visit their general practitioner for any reason [37], including mental health concerns [38]. This difference unsurprisingly extends to adolescents, with the stigma and gender stereotypes surrounding masculinity and strength [39,40] influencing mental health help-seeking likelihood. Interestingly, although female adolescents also tend to be better at identifying psychological problems and possess greater knowledge of available help, education level is also a significant predictor of these factors [41]. Although our finding contributes to the idea that targeted mental health awareness for men and boys is undeniably required, an overall increase in mental health discussion and signposting could help to bridge any educational or socioeconomic differences that exist in terms of mental health literacy and resultant access to support.

#### User-Reported Experiences

More male users than female users chose the domains *important to me* and *understand myself* as reflective of the experience they had with the community post and content, and more female users than male users chose *relate to others*. This suggests that male users may approach web-based community forums with a desire for intrapersonal advice, with female users more likely to seek interpersonal support. Although little research evidence exists to corroborate these findings, they could tentatively match patterns of help-seeking strategies described in a recent study [39], where male participants preferred to adopt self-reliance when dealing with psychological problems. By contrast, female participants tended to have more confidence in mental health professionals. However, the same study found that male participants, more so than female participants, find their friends more helpful under such circumstances [39]; this is a finding that is not concordant with ours. Interestingly, although female adolescents may desire stronger connection from their

friendships, perhaps explaining their increased likelihood of choosing *relate to others*, friendships are valued as equally important by both [42]. In terms of age, we found that older CYP ( $\geq 14$  years) who completed the POCEMs selected *relate to others* more readily than those aged  $\leq 13$  years, with the latter group more likely to choose all 3 of the other experience domains. The association between peer support and well-being becomes stronger for CYP with age [43], which could explain this finding. Friendship, in particular dyadic friendships and the unique social nuances and problems that arise in them, become increasingly salient during early adolescence [44,45].

Figure 1 shows the frequency with which pairs of additional selectable statements were chosen together when users elaborated on the experiences that they had with community content forum posts. We will now discuss some of the key trends that can be inferred from these heat maps. An overall observed pattern was that the statements selected together tended to be conceptually similar; for example, under the *learn skills* domain, *I now have knowledge and skills to help others* and *I have learned how to support others* were chosen as the most frequent pair, both of which are almost identical in terms of the experience reported. Conversely, those least frequently selected together were often clear opposites; for example, under *relate to others*, the items *I know who to ask for help* and *I feel motivated to give advice to others* refer to receiving and providing advice, respectively. Therefore, it is unsurprising that users do not report getting both experiences from a single forum post. Looking more closely at specific trends within each heat map in turn, under *important to me*, the statement *I learned something new today* was the least selected overall, with *I now know that others have the same experiences as me* combined with both *I got information that helped me learn about myself* and *the information I received today was helpful to my problem* representing the most frequent pairings. The latter 2 statements can be viewed as broad, unspecific positive experiences that can easily be combined with others. This point about statement specificity is also relevant for the domain *learn skills*, where the least frequently selected statements *I have learned how to express myself* and *I have developed skills to open up more* can be viewed as less broad than other statements in the domain. They also seemingly relate to personal gain, whereas the others relate more closely to gaining skills to help others. Under *relate to others*, the items *I feel safe in the Kooth community*, *it feels good not to be judged*, and *I felt connected to someone* were commonly selected together, seemingly encompassing a sense of comfort and security. Finally, in *understand myself*, the most frequent pairings, of which *I felt accepted* and *I now feel more hopeful* is 1 example, seem to relate conceptually to positivity and a better appraisal of the users' situation. *I now feel able to ask for support outside of Kooth* was the least commonly selected outcome statement in this domain. It stands out as distinct and not truly relating to the *understand myself* domain, perhaps explaining why it was not chosen if self-understanding was the key overarching experience gained.

## Trends of Forum Post Helpfulness

### Qualities of Forum Posts

In the second data set, significant correlational relationships of varying strengths were found among the qualities of a post at each instance of measure completion. Posts with higher POCEM scores were older, and users spent longer time reading them. Although these correlational relationships were negligible, post age and time spent interacting were key factors of interest, and more POCEMs were also completed on older and longer-read posts. As mentioned earlier, within Kooth's forums, posts do not move to the top of a page when a new comment is posted. In addition, the only way to attempt to locate relevant posts is by selecting a topic category. Accessing and completing the measure for an older post therefore implies a targeted search for that content: the user has likely scrolled through several pages and selected a topic to find information pertaining to their issue. They have likely logged in and spent time engaging with the community because of a specific want or need. Once they find the relevant information, it is likely to be viewed more favorably because of this specific need resolution. We also found that older posts were engaged with for longer before POCEM completion, which further strengthens this idea of targeted searches being more useful—more time is spent reading and absorbing the information within, leading to increased likelihood of POCEM completion and higher rating. Posts interacted with for longer before POCEM completion also had more comments and views. Comments and views are the only 2 engagement figures visible to forum users; therefore, high numbers may elicit a curiosity that results in longer reading time. These relationships all suggest that a more user-friendly search function would be beneficial. Older posts are evidently scrolled through, interacted with, and deemed useful, and the findings suggest that CYP are using the forums to search for information and to seek help in a targeted fashion, in addition to general browsing and interacting. A search-term function, in addition to a facility to sort posts by factors such as *most popular*, would prove useful for users, as well as maximize the research potential of the POCEM itself. In summary, these links among factors such as time spent, post age, and POCEM completion do imply that feedback mechanisms such as the POCEM can be seen not only as proxies of engagement but also as helpful tools for building understanding of web-based forum behavior. More interdisciplinary research of this kind could help us to understand whether engagement and interaction behavior differ depending upon the nature of the web-based system used, be it educational, support seeking, or any other kind. This is due to the fact that our findings do not completely align with studies that looked at engagement in different types of platforms [23,24].

### Time of Day

Looking at POCEM completion across times of day in the web-based community proved insightful, particularly in terms of how CYP choose to spend their time outside of school or work—their *leisure* time. Of the 17,137 POCEMs in data set 2, 10,334 (60.3%) were completed outside of standard working hours (between 4 PM and 6 AM), allowing us to make the broad claim that CYP are actively choosing to engage with community

peer support during their free time. Figures 2 and 3 show that POCEMs were completed for significantly older posts in the morning and afternoon than in the evening, night, and late night. Although this may tenuously suggest that users are searching through older posts during the day, it may simply reflect the fact that higher overall engagement after 4 PM likely includes creation of more new posts at this time. Therefore, if a user logs in between 6 AM and 4 PM, the posts they interact with are likely to be slightly older. This brings to light a limitation that should be considered with this main effect as well as the aforementioned correlational findings: to infer user motivations behind completing an experience measure for older posts, we would need to know the average age and rank of the posts that appear on the first page of the forum when users log in. Although some users in the sample did complete POCEMs for posts that were up to 30 days old, the mean age of a post at POCEM completion was 4.04 (SD 4.21) days. Therefore, interacting with a slightly older post does not necessarily suggest, in all instances, a targeted search. More information about post rate and subsequent first-page turnover is necessary to substantiate this claim. Users also spent longer time engaging with posts before completing the measures in the morning and evening than late at night. Any explanations for this, such as a user having more comments to catch up on in the morning and evening or that more targeted searches late at night perhaps equate to a quicker decision regarding a post's helpfulness, would be purely speculative.

Arguably, the most interesting findings concerning time of day appeared when we examined the topics of posts that were rated at different times of day. As shown in Figure 2, of the posts relating to mental health, 33.61% (656/1952) were completed late at night, whereas 22.87% (739/3231) were completed in the morning. This suggests that users logging into the web-based service at night and in the very early hours are doing so to seek mental health support. This idea is also strengthened when we look at hobbies or interests, a contrasting topic that is more conversational and less advice centric. Significantly more measures on this topic were completed in the morning than at night (Figure 3). How, then, can we explain this seemingly increased focus on mental health at night? Why are those who are awake at night interacting with mental health topics? Insomnia has frequently been linked to an increased likelihood of mental health problems, including in adolescents [46,47]. In addition to insomnia, *eveningness*, often colloquially known as being a *night owl*, is also independently associated with psychopathology in adolescents [11]; however, only insomnia increased the risk of suicide ideation [48]. CYP commonly shift to a pattern of *eveningness* as they enter adolescence, and societal pressures such as school start times do not accommodate this [9,10], perhaps explaining why these CYP have adverse mental health outcomes even in the absence of disordered sleep. In terms of support seeking, patients seeking psychiatric support outside of daytime care hours tend to present with mental health issues of greater severity and complexity [49,50]. The direction of the relationship between poor mental health and being awake at night, either through insomnia or individual differences in sleep patterns, is unclear. However, what is apparent is that users who are awake while most of their peers are asleep are likely to be those with the most complex mental health issues,

which lack of sleep may exacerbate their tendency to ruminate upon at this time [51]. The key suggestion with regard to this finding, although more research is needed to further pinpoint patterns of user activity *after hours*, is to ensure that appropriate provision is in place to support CYP who seek mental health support at this time of day; for example, making sure that round-the-clock crisis helplines such as Nightline (a support line for students in the United Kingdom), The Samaritans, or Childline are especially visible to users after 10 PM when live support is unavailable, with users encouraged to call the helplines if they feel that they require live support urgently.

### Strengths, Limitations, and Future Directions

Although several useful insights were gained from this research, they must be viewed in the context of the following limitations. The first of these relates to the study's cross-sectional design. The large data set of 23,443 POCEM completions was an asset of the study, given the statistical power this offered our analyses. However, examining trends over time by means of a longitudinal design would allow us to explore how forum use and user experiences fluctuate over time. This would be of particular interest during times of widespread crisis, such as the COVID-19 pandemic. This would also allow examination of how long-term forum users engage. Linked to this idea, we mentioned in the Methods section that the 23,443 completed POCEMs came from 11,045 unique users. Given that the study's aims primarily focused on post helpfulness rather than repeated patterns by users, our analyses did not account for the fact that many users completed the measure multiple times. We suggest that each POCEM represents a valid and distinct user experience. However, future longitudinal research could investigate user intensity and patterns of response in 1-time users compared with repeat users. This could be to identify whether helpfulness ratings differ and whether users tend to report the same or different experience domains whenever they interact with the POCEM.

It is also wise to acknowledge the limitations that exist with the finding that 74.6% (8240/11,045) of the POCEM scores were positive. Ratings of satisfaction tend to be completed more frequently by those at opposite ends of the experience spectrum. This extremity bias [52] means that people are very keen to report an excellent or a terrible experience but are less inclined to report an average or uninteresting one. Although these biases are most often reported in consumer research, our finding could easily be explained by this phenomenon, with forum posts perhaps being subject to an even starker imbalance—one that echoes the positively skewed J-shaped distribution curve often reported in web-based product reviews [53]. A forum post, as content nested within the web-based peer community, is likely to either be *very* helpful to a CYP or unhelpful in that it is viewed as either boring or irrelevant. Practitioner moderation means that content in the community forum is unlikely to be actively harmful or damaging, meaning that the only *extreme* experience a CYP can have is a positive one. Therefore, finding a post mediocre is less likely to result in completion of the measure. Encouraging review submission may result in a more even distribution of ratings [52]. Although care must be taken not to pressure users or create an interface that is an annoyance to interact with, increasing the visibility of the measure inside



of the peer community forum may help reduce this bias. This limitation, realistically, applies to all findings in this study when we think about *who* completes a web-based community measure such as this one and *why*. Social desirability and acquiescence biases may also be at play, given the positive skew and given that these bias effects are often seen in digital scale measures [54,55]. However, these effects were raised and addressed in the piloting stage of POCEM development (Mindel, C, unpublished data, November 2021), and consequently, a reassuring message regarding anonymity and confidentiality was added to the measure before the present data were collected. This has hopefully ameliorated any worries that users may have about their responses being linked back to them. It should also be noted that time spent completing the POCEM was included in the recording of time spent interacting with a forum post in data set 2. This time should certainly be separated in future research into the POCEM because it could then be ascertained whether users were reading and engaging with the questions the POCEM asks or whether they were simply clicking random answers. Comparing POCEM completion time with time spent engaging with the related forum post would also prove insightful.

There are also several methodological constraints relating to the use of heat maps (Figure 1). Although heat maps are attractive forms of data visualization, it is important to bear in mind that they are not, and should not be treated as, complete data analyses in their own right [56]. The findings relating to Figure 1 should therefore be treated as *patterns* of engagement rather than true differences owing to the lack of significance testing in heat maps. Along these lines, the colors used when shading heat maps can make differences appear large, even if the gap between the biggest and smallest value is very small.

If they are to be used as analytic tools, measures of central tendency, at the very least, should be included. Heat maps were included in this study for identification of trends and patterns rather than to identify definitive differences. Therefore, provided that these limitations are considered when interpreting the aforementioned findings, they remain a useful way to visualize this element of our data.

## Conclusions

To summarize this research, this study used a large data set to explore the use of a user-rated helpfulness measure in a CYP mental health peer support forum context. The results seem to indicate that the young service users involved in this study found web-based peer support helpful and that the POCEM itself was engaged with well. This suggests that peer support can provide an important strand of care within a supportive mental health ecosystem, particularly during time periods when in-person support is typically closed. This latter finding stresses the importance of vigilant support provision at these times. Measures were most often completed by secondary school-aged CYP and by many more female users than male users. The age of a post and the time spent engaging with a post before POCEM completion were found to be factors of interest, providing insight into the search habits of users. These factors denote key areas where peer support forums can be made more intuitive and user friendly. Caution is advised when interpreting the results of this study. Although such services are popular, and indeed useful, they have received little research attention to date. As such, further investigation into the nature of helpful and unhelpful peer support is warranted, as well as increased focus on how it should be incorporated into comprehensive systems of digital mental health support.

## Acknowledgments

The time spent on this paper by EB, SDOG, CM, TK, and AS was funded by Kooth Digital Health. Special thanks to Holly Brick and Katherine Jackson who helped to develop and conceptualize the measure; to the product team at Kooth for implementing the measure in the digital platform; to Dr Lynne Green and Dr Hannah Wilson, among other members of the clinical team, for safeguarding and overseeing the research project at Kooth; and to Lex Young as well as the content team and moderators for helping to run the forum community.

## Conflicts of Interest

SDOG, CM, TK, and AS are employed by Kooth Digital Health where the data for this study were collected. EB was externally commissioned and funded by Kooth to work on the project but is not an employee of the company.

## References

1. Murphy R, Higgins A. The complex terrain of peer support in mental health: what does it all mean? *J Psychiatr Ment Health Nurs* 2018 Sep;25(7):441-448 [FREE Full text] [doi: [10.1111/jpm.12474](https://doi.org/10.1111/jpm.12474)] [Medline: [29797631](https://pubmed.ncbi.nlm.nih.gov/29797631/)]
2. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. *Soc Sci Med* 2005 Apr;60(7):1467-1478. [doi: [10.1016/j.socscimed.2004.08.010](https://doi.org/10.1016/j.socscimed.2004.08.010)] [Medline: [15652680](https://pubmed.ncbi.nlm.nih.gov/15652680/)]
3. El Asam A, Samara M, Terry P. Problematic internet use and mental health among British children and adolescents. *Addict Behav* 2019 Mar;90:428-436. [doi: [10.1016/j.addbeh.2018.09.007](https://doi.org/10.1016/j.addbeh.2018.09.007)] [Medline: [30579146](https://pubmed.ncbi.nlm.nih.gov/30579146/)]
4. Children and Parents: Media Use and Attitudes Report. Ofcom. 2017. URL: [https://www.ofcom.gov.uk/data/assets/pdf\\_file/0020/108182/children-parents-media-use-attitudes-2017.pdf](https://www.ofcom.gov.uk/data/assets/pdf_file/0020/108182/children-parents-media-use-attitudes-2017.pdf) [accessed 2021-09-27]
5. Hanley T, Prescott J, Gomez KU. A systematic review exploring how young people use online forums for support around mental health issues. *J Ment Health* 2019 Oct;28(5):566-576. [doi: [10.1080/09638237.2019.1630725](https://doi.org/10.1080/09638237.2019.1630725)] [Medline: [31267813](https://pubmed.ncbi.nlm.nih.gov/31267813/)]

6. Hanley T, Sefi A, Grauberg J, Prescott J, Etchebarne A. A theory of change for Web-based therapy and support services for children and young people: collaborative qualitative exploration. *JMIR Pediatr Parent* 2021 Mar 22;4(1):e23193 [FREE Full text] [doi: [10.2196/23193](https://doi.org/10.2196/23193)] [Medline: [33749615](https://pubmed.ncbi.nlm.nih.gov/33749615/)]
7. How Covid-19 is Affecting The Mental Health of Children and Young People. XenZone. 2020. URL: [https://xenzone.com/wp-content/uploads/2020/05/150520\\_CYP\\_infographic\\_D4.pdf](https://xenzone.com/wp-content/uploads/2020/05/150520_CYP_infographic_D4.pdf) [accessed 2021-10-22]
8. Malik S, Coulson NS. The therapeutic potential of the internet: exploring self-help processes in an internet forum for young people with inflammatory bowel disease. *Gastroenterol Nurs* 2011;34(6):439-448. [doi: [10.1097/SGA.0b013e318237a9ba](https://doi.org/10.1097/SGA.0b013e318237a9ba)] [Medline: [22129797](https://pubmed.ncbi.nlm.nih.gov/22129797/)]
9. Carskadon MA, Barker DH. Editorial perspective: adolescents' fragile sleep-shining light on a time of risk to mental health. *J Child Psychol Psychiatry* 2020 Oct;61(10):1058-1060. [doi: [10.1111/jcpp.13275](https://doi.org/10.1111/jcpp.13275)] [Medline: [32496579](https://pubmed.ncbi.nlm.nih.gov/32496579/)]
10. Carskadon MA. Sleep in adolescents: the perfect storm. *Pediatr Clin North Am* 2011 Jun;58(3):637-647 [FREE Full text] [doi: [10.1016/j.pcl.2011.03.003](https://doi.org/10.1016/j.pcl.2011.03.003)] [Medline: [21600346](https://pubmed.ncbi.nlm.nih.gov/21600346/)]
11. Li SX, Chan NY, Man Yu MW, Lam SP, Zhang J, Yan Chan JW, et al. Eveningness chronotype, insomnia symptoms, and emotional and behavioural problems in adolescents. *Sleep Med* 2018 Jul;47:93-99. [doi: [10.1016/j.sleep.2018.03.025](https://doi.org/10.1016/j.sleep.2018.03.025)] [Medline: [29778920](https://pubmed.ncbi.nlm.nih.gov/29778920/)]
12. Prescott J, Hanley T, Ujhelyi Gomez KU. Why do young people use online forums for mental health and emotional support? Benefits and challenges. *Br J Guid Couns* 2019 May 20;47(3):317-327 [FREE Full text] [doi: [10.1080/03069885.2019.1619169](https://doi.org/10.1080/03069885.2019.1619169)]
13. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. *JMIR Ment Health* 2015 May 19;2(2):e19 [FREE Full text] [doi: [10.2196/mental.4418](https://doi.org/10.2196/mental.4418)] [Medline: [26543923](https://pubmed.ncbi.nlm.nih.gov/26543923/)]
14. Prescott J, Hanley T, Ujhelyi K. Peer communication in online mental health forums for young people: directional and nondirectional support. *JMIR Ment Health* 2017 Aug 02;4(3):e29 [FREE Full text] [doi: [10.2196/mental.6921](https://doi.org/10.2196/mental.6921)] [Medline: [28768607](https://pubmed.ncbi.nlm.nih.gov/28768607/)]
15. Lasanen M, Määttä K, Uusiutu S. 'I am not alone' – an ethnographic research on the peer support among northern-Finnish children with hearing loss. *Early Child Dev Care* 2019;189(7):1203-1218 [FREE Full text] [doi: [10.1080/03004430.2017.1371704](https://doi.org/10.1080/03004430.2017.1371704)]
16. Horgan A, McCarthy G, Sweeney J. An evaluation of an online peer support forum for university students with depressive symptoms. *Arch Psychiatr Nurs* 2013 Apr;27(2):84-89. [doi: [10.1016/j.apnu.2012.12.005](https://doi.org/10.1016/j.apnu.2012.12.005)] [Medline: [23540518](https://pubmed.ncbi.nlm.nih.gov/23540518/)]
17. Smithson J, Sharkey S, Hewis E, Jones R, Emmens T, Ford T, et al. Problem presentation and responses on an online forum for young people who self-harm. *Discourse Stud* 2011 Jul 27;13(4):487-501 [FREE Full text] [doi: [10.1177/1461445611403356](https://doi.org/10.1177/1461445611403356)]
18. Frost J, Vermeulen IE, Beekers N. Anonymity versus privacy: selective information sharing in online cancer communities. *J Med Internet Res* 2014 May 14;16(5):e126 [FREE Full text] [doi: [10.2196/jmir.2684](https://doi.org/10.2196/jmir.2684)] [Medline: [24828114](https://pubmed.ncbi.nlm.nih.gov/24828114/)]
19. Moore MJ, Nakano T, Enomoto A, Suda T. Anonymity and roles associated with aggressive posts in an online forum. *Comput Human Behav* 2012 May;28(3):861-867 [FREE Full text] [doi: [10.1016/j.chb.2011.12.005](https://doi.org/10.1016/j.chb.2011.12.005)]
20. Paskuda M, Lewkowicz M. Anonymity interacting with participation on a Q&A site. *AI Soc* 2017;32(3):369-381 [FREE Full text] [doi: [10.1007/s00146-016-0660-9](https://doi.org/10.1007/s00146-016-0660-9)]
21. Lee JT, Yang MC, Rim HC. Discovering high-quality threaded discussions in online forums. *J Comput Sci Technol* 2014 May 17;29(3):519-531 [FREE Full text] [doi: [10.1007/s11390-014-1446-5](https://doi.org/10.1007/s11390-014-1446-5)]
22. Chaka C, Nkhobo T. Online module login data as a proxy measure of student engagement: the case of myUnisa, MoyaMA, Flipgrid, and Gephi at an ODeL institution in South Africa. *Int J Educ Technol High Educ* 2019 Oct 04;16(1):38 [FREE Full text] [doi: [10.1186/s41239-019-0167-9](https://doi.org/10.1186/s41239-019-0167-9)]
23. Henrie CR, Bodily R, Larsen R, Graham CR. Exploring the potential of LMS log data as a proxy measure of student engagement. *J Comput High Educ* 2018;30(2):344-362 [FREE Full text] [doi: [10.1007/s12528-017-9161-1](https://doi.org/10.1007/s12528-017-9161-1)]
24. Ober TM, Hong MR, Rebouças-Ju DA, Carter MF, Liu C, Cheng Y. Linking self-report and process data to performance as measured by different assessment types. *Comput Educ* 2021 Jul;167:104188 [FREE Full text] [doi: [10.1016/j.compedu.2021.104188](https://doi.org/10.1016/j.compedu.2021.104188)]
25. Kesmodel US. Cross-sectional studies - what are they good for? *Acta Obstet Gynecol Scand* 2018 Apr;97(4):388-393 [FREE Full text] [doi: [10.1111/aogs.13331](https://doi.org/10.1111/aogs.13331)] [Medline: [29453895](https://pubmed.ncbi.nlm.nih.gov/29453895/)]
26. Noble J, de Ossorno Garcia S, Gadd L, Gillman A. Theory of change Kooth for adults: how adults use the 'positive virtual ecosystem'. Kooth Plc and NPC. 2021 Feb. URL: <https://tinyurl.com/2atmyych> [accessed 2021-09-27]
27. Mc Cord KA, Ewald H, Agarwal A, Glinz D, Aghlmandi S, Ioannidis JP, et al. Treatment effects in randomised trials using routinely collected data for outcome assessment versus traditional trials: meta-research study. *BMJ* 2021 Mar 03;372:n450 [FREE Full text] [doi: [10.1136/bmj.n450](https://doi.org/10.1136/bmj.n450)] [Medline: [33658187](https://pubmed.ncbi.nlm.nih.gov/33658187/)]
28. University ethical approval. The University of Manchester. URL: <https://www.manchester.ac.uk/research/environment/governance/ethics/approval/> [accessed 2022-01-10]
29. BPS Code of Human Research Ethics. British Psychological Society. 2021. URL: <https://www.bps.org.uk/guideline/bps-code-human-research-ethics-0> [accessed 2021-10-08]

30. Hanley T, Sefi A, Grauberg J, Green L, Prescott J. A positive virtual ecosystem: the theory of change for Kooth. Kooth and University of Manchester. 2020 Nov. URL: [https://explore.kooth.com/wp-content/uploads/2021/02/TOC-2020-report-rebrand\\_Full-report\\_digital.pdf](https://explore.kooth.com/wp-content/uploads/2021/02/TOC-2020-report-rebrand_Full-report_digital.pdf) [accessed 2021-09-27]
31. Delacre M, Leys C, Mora YL, Lakens D. Taking parametric assumptions seriously: arguments for the use of Welch's F-test instead of the classical F-test in One-Way ANOVA. *Int Rev Soc Psychol* 2019;32(1):13 [FREE Full text] [doi: [10.31234/osf.io/wnezg](https://doi.org/10.31234/osf.io/wnezg)]
32. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol Psychiatry* 2022 Jan;27(1):281-295 [FREE Full text] [doi: [10.1038/s41380-021-01161-7](https://doi.org/10.1038/s41380-021-01161-7)] [Medline: [34079068](https://pubmed.ncbi.nlm.nih.gov/34079068/)]
33. Wilkinson L, Friendly M. The history of the cluster heat map. *Am Stat* 2009 May;63(2):179-184 [FREE Full text] [doi: [10.1198/tas.2009.0033](https://doi.org/10.1198/tas.2009.0033)]
34. Census. Office for National Statistics. 2020. URL: <https://www.ons.gov.uk/census> [accessed 2021-10-22]
35. Patalay P, Fitzsimons E. Development and predictors of mental ill-health and wellbeing from childhood to adolescence. *Soc Psychiatry Psychiatr Epidemiol* 2018 Dec;53(12):1311-1323. [doi: [10.1007/s00127-018-1604-0](https://doi.org/10.1007/s00127-018-1604-0)] [Medline: [30259056](https://pubmed.ncbi.nlm.nih.gov/30259056/)]
36. Jones PB. Adult mental health disorders and their age at onset. *Br J Psychiatry Suppl* 2013 Jan;54:s5-10. [doi: [10.1192/bjp.bp.112.119164](https://doi.org/10.1192/bjp.bp.112.119164)] [Medline: [23288502](https://pubmed.ncbi.nlm.nih.gov/23288502/)]
37. Thompson AE, Anisimowicz Y, Miedema B, Hogg W, Wodchis WP, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. *BMC Fam Pract* 2016 Mar 31;17:38 [FREE Full text] [doi: [10.1186/s12875-016-0440-0](https://doi.org/10.1186/s12875-016-0440-0)] [Medline: [27036116](https://pubmed.ncbi.nlm.nih.gov/27036116/)]
38. Harris MG, Baxter AJ, Reavley N, Diminic S, Pirkis J, Whiteford HA. Gender-related patterns and determinants of recent help-seeking for past-year affective, anxiety and substance use disorders: findings from a national epidemiological survey. *Epidemiol Psychiatr Sci* 2016 Dec;25(6):548-561 [FREE Full text] [doi: [10.1017/S2045796015000876](https://doi.org/10.1017/S2045796015000876)] [Medline: [26428069](https://pubmed.ncbi.nlm.nih.gov/26428069/)]
39. Bosco N, Giaccherini S, Meringolo P. A gender perspective about young people's seeking help. *J Prev Interv Community* 2020;48(2):132-146. [doi: [10.1080/10852352.2019.1624353](https://doi.org/10.1080/10852352.2019.1624353)] [Medline: [31215327](https://pubmed.ncbi.nlm.nih.gov/31215327/)]
40. Juvrud J, Rennels JL. "I don't need help": gender differences in how gender stereotypes predict help-seeking. *Sex Roles* 2017;76(1-2):27-39 [FREE Full text] [doi: [10.1007/s11199-016-0653-7](https://doi.org/10.1007/s11199-016-0653-7)]
41. Haavik L, Joa I, Hatloy K, Stain H, Langeveld J. Help seeking for mental health problems in an adolescent population: the effect of gender. *J Ment Health* 2019 Oct;28(5):467-474. [doi: [10.1080/09638237.2017.1340630](https://doi.org/10.1080/09638237.2017.1340630)] [Medline: [28719230](https://pubmed.ncbi.nlm.nih.gov/28719230/)]
42. Zarbatany L, Conley R, Pepper S. Personality and gender differences in friendship needs and experiences in preadolescence and young adulthood. *Int J Behav Dev* 2004 Jul 1;28(4):299-310 [FREE Full text] [doi: [10.1080/01650250344000514](https://doi.org/10.1080/01650250344000514)]
43. Chu PS, Saucier DA, Hafner E. Meta-analysis of the relationships between social support and well-being in children and adolescents. *J Soc Clin Psychol* 2010 Jun;29(6):624-645 [FREE Full text] [doi: [10.1521/jscp.2010.29.6.624](https://doi.org/10.1521/jscp.2010.29.6.624)]
44. Berndt TJ. The features and effects of friendship in early adolescence. *Child Dev* 1982 Dec;53(6):1447-1460 [FREE Full text] [doi: [10.2307/1130071](https://doi.org/10.2307/1130071)]
45. Lansford JE, Yu T, Pettit GS, Bates JE, Dodge KA. Pathways of peer relationships from childhood to young adulthood. *J Appl Dev Psychol* 2014 Mar;35(2):111-117 [FREE Full text] [doi: [10.1016/j.appdev.2013.12.002](https://doi.org/10.1016/j.appdev.2013.12.002)] [Medline: [24948845](https://pubmed.ncbi.nlm.nih.gov/24948845/)]
46. Chung KF, Kan KK, Yeung WF. Insomnia in adolescents: prevalence, help-seeking behaviors, and types of interventions. *Child Adolesc Ment Health* 2014 Feb;19(1):57-63. [doi: [10.1111/camh.12009](https://doi.org/10.1111/camh.12009)] [Medline: [32878360](https://pubmed.ncbi.nlm.nih.gov/32878360/)]
47. Hysing M, Harvey A, Bøe T, Heradstveit O, Vedaa Ø, Sivertsen B. Trajectories of sleep problems from adolescence to adulthood. Linking two population-based studies from Norway. *Sleep Med* 2020 Nov;75:411-417. [doi: [10.1016/j.sleep.2020.08.035](https://doi.org/10.1016/j.sleep.2020.08.035)] [Medline: [32971382](https://pubmed.ncbi.nlm.nih.gov/32971382/)]
48. Chen SJ, Zhang JH, Li SX, Tsang CC, Chan KC, Au CT, et al. The trajectories and associations of eveningness and insomnia with daytime sleepiness, depression and suicidal ideation in adolescents: a 3-year longitudinal study. *J Affect Disord* 2021 Nov 01;294:533-542. [doi: [10.1016/j.jad.2021.07.033](https://doi.org/10.1016/j.jad.2021.07.033)] [Medline: [34330050](https://pubmed.ncbi.nlm.nih.gov/34330050/)]
49. Johansen IH, Morken T, Hunskaar S. Contacts related to mental illness and substance abuse in primary health care: a cross-sectional study comparing patients' use of daytime versus out-of-hours primary care in Norway. *Scand J Prim Health Care* 2010 Sep;28(3):160-165 [FREE Full text] [doi: [10.3109/02813432.2010.493310](https://doi.org/10.3109/02813432.2010.493310)] [Medline: [20575649](https://pubmed.ncbi.nlm.nih.gov/20575649/)]
50. Johansen IH, Carlsen B, Hunskaar S. Psychiatry out-of-hours: a focus group study of GPs' experiences in Norwegian casualty clinics. *BMC Health Serv Res* 2011 May 27;11:132 [FREE Full text] [doi: [10.1186/1472-6963-11-132](https://doi.org/10.1186/1472-6963-11-132)] [Medline: [21619608](https://pubmed.ncbi.nlm.nih.gov/21619608/)]
51. Malmberg M, Larsen JK. Depressive symptoms: the interaction between rumination and self-reported insomnia. *Depress Res Treat* 2015;2015:150828 [FREE Full text] [doi: [10.1155/2015/150828](https://doi.org/10.1155/2015/150828)] [Medline: [26617999](https://pubmed.ncbi.nlm.nih.gov/26617999/)]
52. Karaman H. Online review solicitations reduce extremity bias in online review distributions and increase their representativeness. *Manag Sci* 2021 Jul;67(7):4420-4445 [FREE Full text] [doi: [10.1287/mnsc.2020.3758](https://doi.org/10.1287/mnsc.2020.3758)]
53. Hu N, Pavlou PA, Zhang J. On self-selection biases in online product reviews. *MIS Q* 2017 Feb 1;41(2):449-471 [FREE Full text] [doi: [10.25300/misq/2017/41.2.06](https://doi.org/10.25300/misq/2017/41.2.06)]
54. Kuru O, Pasek J. Improving social media measurement in surveys: avoiding acquiescence bias in Facebook research. *Comput Human Behav* 2016 Apr;57:82-92 [FREE Full text] [doi: [10.1016/j.chb.2015.12.008](https://doi.org/10.1016/j.chb.2015.12.008)]

55. Vesely S, Klöckner CA. Social desirability in environmental psychology research: three meta-analyses. *Front Psychol* 2020 Jul 24;11:1395 [FREE Full text] [doi: [10.3389/fpsyg.2020.01395](https://doi.org/10.3389/fpsyg.2020.01395)] [Medline: [32793022](https://pubmed.ncbi.nlm.nih.gov/32793022/)]
56. Bojko A. Informative or misleading? Heatmaps deconstructed. In: *Proceedings of the 13th International Conference on Human-Computer Interaction. New Trends. 2009 Presented at: HCI '09; July 19-24, 2009; San Diego, CA, USA p. 30-39* URL: [https://doi.org/10.1007/978-3-642-02574-7\\_4](https://doi.org/10.1007/978-3-642-02574-7_4) [doi: [10.1007/978-3-642-02574-7\\_4](https://doi.org/10.1007/978-3-642-02574-7_4)]

## Abbreviations

**CYP:** children and young people

**POCEM:** Peer Online Community Experience Measure

*Edited by A Mavragani; submitted 14.01.22; peer-reviewed by D Smit, SJ Lee, J Prescott; comments to author 26.07.22; revised version received 07.08.22; accepted 10.08.22; published 09.09.22.*

*Please cite as:*

*Banwell E, Hanley T, De Ossorno Garcia S, Mindel C, Kayll T, Sefi A*

*The Helpfulness of Web-Based Mental Health and Well-being Forums for Providing Peer Support for Young People: Cross-sectional Exploration*

*JMIR Form Res* 2022;6(9):e36432

URL: <https://formative.jmir.org/2022/9/e36432>

doi: [10.2196/36432](https://doi.org/10.2196/36432)

PMID: [36083629](https://pubmed.ncbi.nlm.nih.gov/36083629/)

©Emily Banwell, Terry Hanley, Santiago De Ossorno Garcia, Charlotte Mindel, Thomas Kayll, Aaron Sefi. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 09.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Quality of Late-Life Depression Information on the Internet: Website Evaluation Study

Teaghan A M Pryor<sup>1\*</sup>, BA; Kristin A Reynolds<sup>1,2\*</sup>, BA, MA, PhD; Paige L Kirby<sup>1\*</sup>, BSc, MOT; Matthew T Bernstein<sup>1\*</sup>, BA, MA, PhD

<sup>1</sup>Department of Psychology, University of Manitoba, Winnipeg, MB, Canada

<sup>2</sup>Department of Psychiatry, University of Manitoba, Winnipeg, MB, Canada

\*all authors contributed equally

**Corresponding Author:**

Kristin A Reynolds, BA, MA, PhD

Department of Psychology

University of Manitoba

190 Dysart Rd

Winnipeg, MB, R3T 2N2

Canada

Phone: 1 2044749528

Email: [kristin.reynolds@umanitoba.ca](mailto:kristin.reynolds@umanitoba.ca)

## Abstract

**Background:** The internet can increase the accessibility of mental health information and improve the mental health literacy of older adults. The quality of mental health information on the internet can be inaccurate or biased, leading to misinformation.

**Objective:** This study aims to evaluate the quality, usability, and readability of websites providing information concerning depression in later life.

**Methods:** Websites were identified through a Google search and evaluated by assessing quality (DISCERN), usability (Patient Education Materials Assessment Tool), and readability (Simple Measure of Gobbledygook).

**Results:** The overall quality of late-life depression websites (N=19) was adequate, and the usability and readability were poor. No significant relationship was found between the quality and readability of the websites.

**Conclusions:** The websites can be improved by enhancing information quality, usability, and readability related to late-life depression. The use of high-quality websites may improve mental health literacy and shared treatment decision-making for older adults.

(*JMIR Form Res* 2022;6(9):e36177) doi:[10.2196/36177](https://doi.org/10.2196/36177)

**KEYWORDS**

late-life; depression; older adults; internet; websites; information quality; usability; readability

## Introduction

**Background**

Late-life depression (LLD) can occur in adults aged  $\geq 65$  years, either for the first time or as a recurrent episode [1]. Major depressive disorder (MDD) is characterized by low mood or loss of interest in daily activities, changes in weight or appetite, trouble sleeping or sleeping too much, lack of energy, feelings of worthlessness or guilt, psychomotor agitation or retardation, difficulty focusing or making decisions, and thoughts of death or suicide. At least five of these symptoms must occur for most of the day, nearly every day, or for a period of at least 2 weeks [2].

Approximately 2%-6% of adults aged  $\geq 55$  years have received a diagnosis of MDD or experienced a major depressive episode within the past year [3-5]. The prevalence of MDD in older adults may be higher when including individuals experiencing subclinical depression [6,7]. Furthermore, the severity of depressive symptoms among older adults is particularly concerning, given that adults aged  $\geq 70$  years, and older men, in particular, have the highest rates of completed suicide worldwide [8].

It is important to note that older adults experiencing LLD can differ in symptomology compared with younger adults, such as presenting with fewer affective symptoms (eg, tearfulness), increased complaints of somatic symptoms, cognitive changes,

and loss of interest [9]. Older adults are also more likely to experience comorbid health conditions and neurological disorders, which further affects the identification of LLD and influences the need for specialized treatment approaches [1,9].

When older adults experience a mental health problem, such as depression, they are faced with a lack of information regarding the symptoms and how to manage them with effective treatment options [10,11]. This gap in knowledge leads to lower levels of mental health literacy (knowledge about recognition, prevention, and management), which can complicate or delay the mental health treatment-seeking process [12,13]. Despite such barriers, most older adults have positive feelings about seeking help for mental health problems and a desire for increased information and participation in the treatment decision-making process [14,15].

Shared decision-making is a process that occurs between a patient and a health care provider, where diverse treatment options are shared by both parties to foster the agreement and implementation of a preferred treatment option [16]. Engaging in this process is beneficial for individuals with mental health problems, increasing their satisfaction and involvement in treatment decisions [17,18].

The internet can be a valuable tool for meeting the information needs of older adults. By presenting a wide variety of treatment options, web-based information can facilitate engagement in the shared treatment decision-making process [19]. As noted previously, web-based depression information should reflect the differences in symptom presentation and treatment options that are relevant to the unique needs of the older adult population [1].

Most internet users (58%-78%) use the internet to search for health information [20-25], and it is increasingly being used to access mental health information [26-29]. Older adults have high rates of internet use, with 73% of older adults using the internet [30,31] and 40% using the internet to access health information [32]. Little is known about specific internet use by older adults for mental health information, although a study found that 11% of older adults used it for finding information on mental health problems and expressed interest in using the internet as a tool for managing their mental health [33]. A more recent study found that 67% of adults with bipolar disorder aged  $\geq 60$  years who used the internet used it to access information about their disorder [34]. Despite this, some older adults feel as though they lack the knowledge and confidence to use the internet as a source of information [35].

Therefore, caregivers of older adults are often involved in seeking information for their care recipient and play an invaluable role in facilitating the treatment-seeking and shared decision-making process [36,37]. Part of the information-seeking process involves using the internet to access essential treatment information. Recent research has shown that a high percentage of caregivers (67%-71%) use the internet to access health information on behalf of the individual they support [38,39], particularly for older adults [40].

Despite the benefits of internet use for health information queries, including anonymity and accessibility, there are some

disadvantages [41]. One of the drawbacks is the uncertainty of the quality of the information provided on the internet [31,42]. It may be difficult for individuals to determine whether the information presented is unbiased, accurate, and evidence based [43-46]. Inaccurate health information retrieved from the internet, which patients incorporate into their clinical requests, has been demonstrated to harm the physician-patient relationship and have detrimental effects on their health outcomes [47]. Therefore, evaluating the quality of the information provided on the web is essential in preventing the spread of misinformation and facilitating increased knowledge of balanced treatment options not only for older adults themselves but also for those who seek out information on their behalf.

As a result, studies evaluating website quality are increasing [48] as internet use for mental health information becomes more prevalent. Furthermore, studies have raised concerns regarding the approach that researchers have taken to evaluate websites and the lack of consistency across studies. An important question stemming from this growing research base is what constitutes a high-quality website. Common criteria used in the literature to identify high-quality websites are based on the principles of quality, usability, and readability [48-50]. Website quality involves the extent to which a website provides clear, unbiased, evidence-based information regarding mental health diagnoses and treatment options [51,52]. Usability characteristics refer to the ease of use, navigation, and aesthetics of websites [53], and readability is the ease of reading written text [54-56], both of which contribute to the overall quality of web-based information. Therefore, high-quality websites should be relatively straightforward to navigate and comprehend.

Website evaluations have been completed on a variety of health topics, including depression in the adult population [57-61]. A systematic review of studies evaluating the quality of mental health websites found that 23 of the 31 studies had poor quality overall [48]. A small number of studies have been conducted assessing the quality of depression information provided by websites. Generally, studies have shown that the quality of website information is poor [57-59]. However, one study examining the overall quality of websites for depression in adults aged 18 to 64 years found adequate quality, with most websites scoring higher than the mean score on a measure of content quality [61]. The discrepancy in findings could be because of variability in the website evaluation methodology used across studies. Nevertheless, this range of quality (poor to adequate) may not be helpful to consumers.

## Objectives

Upon extensive review of the website evaluation literature, to the best of our knowledge, no study has examined the quality of websites specific to depression that appears or worsens in later life (LLD). Given the unique impact of aging on depression presentation, the involvement of caregivers in the treatment-seeking processes, and the influence of web-based information in the shared decision-making process, we deemed it essential to evaluate the quality of websites providing information on LLD. Websites were evaluated according to (1) quality of information, as evaluated by DISCERN [62], a standardized measure of website quality; (2) usability, as

determined by the Patient Education Materials Assessment Tool (PEMAT) [63]; and (3) readability of information, as evaluated by the Simple Measure of Gobbledygook (SMOG) [64]. A secondary objective of this study was to determine whether website quality was related to usability and readability.

On the basis of the existing mental health website evaluation literature, we hypothesized that (1) website quality would be adequate to poor according to DISCERN evaluations, (2) usability would be adequate to poor, and (3) the reading level would be higher than the recommended levels for health information (grades 7-8). Furthermore, we hypothesized that website quality, usability, and readability would be related. Specifically, the quality of websites would be positively associated with usability scores and negatively associated with reading levels.

## Methods

### Website Selection

Google Canada search engine was used to identify websites as it is the most widely used search engine worldwide [65] and has been identified as the starting point for most internet users seeking health information [30,42,66]. The search was completed on one of the computers in the research laboratory by the primary investigator (TAMP), where website cookies and search history were cleared before searching to prevent influencing the search results. The search terms “older” AND “depression” were used to complete the initial search to target websites presenting distinct information on LLD in this age group (presentation and treatment options).

Websites within the first 3 pages of the search were evaluated if they did not meet exclusion criteria, as it has been found that most search engine users do not go past the first 3 pages of the search [67]. Websites were excluded from the evaluation if they were advertising or selling products; presented information from books or articles, as the purpose of this study is to evaluate websites; contained minimal information (<500 words) that was not substantial enough to evaluate; provided information that was not focused on LLD; and were not written in English, as this is the researchers’ language of origin. The website selection method and exclusion criteria were in line with prior depression website evaluation research [57-60].

### Procedure

#### Measures of Website Quality

The DISCERN instrument is a standardized measure comprising 16 items assessing the quality of written health information [62] and has been used in a variety of health website evaluation studies [68-71]. The reliability of DISCERN has been psychometrically evaluated in previous research [72-74] and is able to differentiate between low- and high-quality information [74]. DISCERN comprises 3 main sections focusing on how reliable the publication is, the quality of information for treatment options, and the overall quality of the publication [62]. Each item is rated on a 5-point scale from no to yes indicating the extent to which the information fulfills the criteria:

1 (criterion was not fulfilled at all), 2 to 4 (criterion was fulfilled to some extent), and 5 (criterion was completely fulfilled).

#### Usability

The PEMAT was used to assess usability. The PEMAT is a multi-item standardized tool used to assess the understandability (ability to understand materials) and actionability (ability to encourage consumers to take action on information presented) of materials educating patients on a variety of health topics [63]. The PEMAT has been used in numerous recent studies evaluating printed and web-based health-related materials [75,76]. The tool has been evaluated and found to have good internal consistency, reliability, and construct validity [77,78]. The tool is divided into 2 domains—understandability and actionability—with specific topic areas under each domain. The number of items used varies depending on the type of material used, either printed or audiovisual. For this study, PEMAT-printed was used, which comprises 34 items (17 understandability and 7 actionability items), as most of the website information can be printed from each website. Information was rated according to each item and scored either 0=*disagree* or 1=*agree*, with some items having the option of *not applicable*=NA. Separate usability and actionability scores were calculated by summing the total number of points given (excluding *not applicable* items) and dividing by the total number of possible points. This number was multiplied by 100 to obtain a percentage score to determine what percentage of the material is understandable or actionable.

#### Readability

A readability score was calculated for each website using the SMOG. The SMOG assesses the number of words with  $\geq 3$  syllables in 10 consecutive sentences sampled from the beginning, middle, and end of the text [79]. Although the Flesh-Kinkaid reading formula has been most frequently used and cited in assessing the readability of health information, the SMOG formula is recommended as a more appropriate formula to assess health information because of its consistency and ease of use when calculating reading level [79].

#### Analyses

The analysis component comprised (1) descriptive, (2) correlational, and (3) inferential statistics. All statistics were computed by the first author (TAMP) using SPSS (version 21.0; IBM Corp) for Windows. Descriptive analyses produced a mean score for each website, as well as a mean score and 95% CI for each DISCERN item. In addition, 2 percentages were computed for the domains of understandability and actionability according to the PEMAT. A 2-tailed Pearson correlation was calculated to determine the relationship between website quality (as determined by DISCERN) and website usability scores (as determined by the PEMAT) and between website quality and website reading level independently. An intraclass correlation coefficient was computed between the first half of the selected websites to determine the level of agreement between the primary (TAMP) and secondary (PLK) raters for DISCERN and the PEMAT, similar to previous research [60,80].

## Ethical Considerations

This research did not involve human participants and involved the examination of mental health information in the public domain. Therefore, it was determined that no ethics approval was required to carry out this study.

## Results

### Website Characteristics

Evaluated websites (N=19) are described in [Table 1](#). This sample is consistent with previous research [57-60]. Most of the

websites were from the United States; however, websites from Australia (Beyond Blue), Canada (HealthLinkBC), and Great Britain (Royal College of Psychiatry) also emerged in the initial search. The evaluated websites were hosted by hospitals (eg, Johns Hopkins Medicine), nonprofit organizations (Age UK and HelpGuide), and government organizations (National Institute of Mental Health). Most of the websites did not have a formal operationalized definition of LLD and used a variety of terms when defining LLD (eg, elderly or geriatric depression, depression and older adults or in older people, and aging and depression). Websites identified the ages of 60 to ≥65 years as a target age group within the late-life period.

**Table 1.** Late-life depression website characteristics.

Website (country)	Search engine order	Overall quality <sup>a,b</sup>	Usability <sup>c,d</sup>		Readability <sup>e</sup>
			Understandability score (%)	Actionability score (%)	
Age UK (United Kingdom)	24	2.9	69.2	60	9.5
American Psychological Association (United States)	12	2.3	53.8	40	11.1
Beyond Blue (Australia)	4	4.1	53.8	60	12.5
Black Dog Institute (Australia)	16	2.5	76.9	20	12.4
Canadian Coalition for Senior's Mental Health (Canada)	28	3.0	84.6	60	6.9
Centers for Disease Control and Prevention (United States)	10	2.3	46.2	40	10.7
Health in Aging (United States)	17	4.0	53.8	60	11.5
Healthline (United States)	11	3.6	76.9	60	10.8
HealthLinkBC (Canada)	30	3.9	76.9	80	7.4
HelpGuide (United States)	1	3.4	71.4	60	9.6
Johns Hopkins Medicine (United States)	13	2.6	69.2	40	10.9
MedlinePlus (United States)	9	2.3	66.7	40	8.8
Mental Health America (United States)	5	3.4	61.5	80	11.2
National Institute of Mental Health (United States)	3	3.9	75.0	40	10.2
National Institute on Aging (United States)	2	3.3	53.8	60	10.6
Royal College of Psychiatrists (United Kingdom)	8	4.8	69.2	50	7.1
Substance Abuse and Mental Health Services (United States)	29	3.9	57.1	20	12.4
WebMD (United States)	6	3.6	46.2	40	9.8
World Health Organization	25	1.5	30.8	0	13.5

<sup>a</sup>Overall quality was measured by DISCERN.

<sup>b</sup>The mean DISCERN score is a 1 to 5 rating averaged across 16 items.

<sup>c</sup>Usability was measured by the Patient Education Materials Assessment Tool.

<sup>d</sup>To calculate the scores, items that are agreed upon are summed and divided by the total possible points then multiplied by 100 to get a percentage.

<sup>e</sup>Readability was measured by the Simple Measure of Gobbledygook.



## Website Quality

A mean score was provided for each website to better understand the quality of the information presented. Website quality varied greatly, with scores ranging from 1.5 (low quality) to 4.8 (high quality) out of a total score of 5 (Table 1). Websites that scored highly on DISCERN included *the Royal College of Psychiatrists* (4.8), *Beyond Blue* (4.1), and *Health in Aging* (4.0).

An average score across the websites was also computed for each DISCERN item to better understand the criteria that were well addressed and criteria that needed improvement (Table 2). Most of the websites addressed the DISCERN items moderately well, with average DISCERN scores ranging from 2.4 to 4.0 (SD 1.0 to 1.8). The mean score of item 16, which served as an overall rating of the websites, was 3.2. This indicates that

websites with information about LLD were of adequate quality. According to scores on specific DISCERN criteria, websites clearly showed that there were multiple treatment options available (average score of 4.0). Information presented by websites was relevant to the older adult population (average score of 3.7). Websites also encouraged shared decision-making with health care providers or family members (average score of 3.7). Websites lacked information on the risks of treatment options (average score of 2.4), describing how each treatment worked (average score of 2.5), and the benefits of each treatment (average score of 2.6). Most websites did not provide the sources used to create the publication (average score of 2.6). An intraclass correlation was computed to determine the reliability of the raters on the DISCERN measure, which determined that there was an excellent level of agreement ( $r_g=0.90$ ;  $P<.001$ ).

**Table 2.** Mean scores of DISCERN items across all websites<sup>a</sup>.

Item number	DISCERN item	Score, mean (SD; 95% CI)
1	Are the aims clear?	3.4 (1.1; 2.9-3.9)
2	Does it achieve its aims?	3.6 (1.0; 3.2-4.1)
3	Is it relevant?	3.7 (1.2; 3.2-4.2)
4	Is it clear what sources of information were used to compile the publication (other than the author or producer)?	2.6 (1.6; 1.9-3.3)
5	Is it clear when the information used or reported in the publication was produced?	3.3 (1.3; 2.7-3.9)
6	Is it balanced and unbiased?	3.6 (1.0; 3.1-4.0)
7	Does it provide details of additional sources of support and information?	3.2 (1.4; 2.6-3.9)
8	Does it refer to areas of uncertainty?	3.1 (1.3; 2.5-3.7)
9	Does it describe how each treatment works?	2.5 (1.2; 2.0-3.1)
10	Does it describe the benefits of each treatment?	2.6 (1.2; 2.1-3.2)
11	Does it describe the risks of each treatment?	2.4 (1.3; 1.8-3.0)
12	Does it describe what would happen if no treatment was used?	2.8 (1.8; 2.0-3.6)
13	Does it describe how the treatment choices affect the overall quality of life?	3.4 (1.0; 3.0-3.9)
14	Is it clear that there may be more than one possible treatment choice?	4 (1.3; 3.4-4.6)
15	Does it provide support for shared decision-making?	3.7 (1.2; 3.2-4.3)
16	Based on the answers to all of the above questions, rate the overall quality of the publication as a source of information about treatment choices.	3.2 (1.2; 2.6-3.7)

<sup>a</sup>Each DISCERN item is rated on a 5-point scale with the anchors 1=did not meet criteria and 5=did meet criteria.

## Usability

The understandability scores of the PEMAT (Table 1) varied, ranging from 30.8% to 84.6% (mean 62.8%). Only 32% (6/19) of websites met the 70% criteria, indicating that the website was understandable [77]. With regard to specific understandability criteria, most websites used an active voice and used a variety of visual cues to draw attention to important points of the websites. Websites lacked summaries of information and introduced complex medical terms within the text without definition. In examining the actionability section of the PEMAT, websites presented a range of scores from 0% to 80% (mean 47.9%), with only 10% (2/19) of the 19 websites meeting the minimum 70% threshold for websites to be deemed actionable. Upon further examination, most websites identified

at least one action that individuals could take (eg, talking to their physician). Despite this, most websites did not provide any visual aids encouraging individuals to take action, lacked tools to aid individuals in taking action (eg, treatment planning sheet), and did not break down suggested actions into explicit steps.

Two separate correlations were computed between the mean DISCERN scores and the usability and actionability percentages of the PEMAT to determine the relationship between website quality and usability. The correlation between DISCERN scores and the understandability scores of the PEMAT was not significant ( $r_{17}=-0.30$ ;  $P=.21$ ). By contrast, the correlation between the DISCERN scores and the actionability scores of the PEMAT was found to be significant ( $r_{17}=0.49$ ;  $P=.04$ ). An

intraclass correlation was also computed to determine the interrater reliability, which established that there was an excellent level of agreement for the understandability section ( $r_g=0.90$ ;  $P<.001$ ).

### Readability

The readability of the websites was calculated using the SMOG readability formula, which produced a grade level score. The reading levels of the websites ranged from 6.9 to 13.5, with an average grade level of 10.4 across all websites (Table 1). Only 16% (3/19) of the websites met the National Institute of Health's recommended grade level (grade 7-8). A correlational analysis was conducted to determine whether website readability was related to website quality, as measured by DISCERN. This analysis was nonsignificant ( $r_{17}=-0.31$ ;  $P=.20$ ).

### Website Dimension Comparison

Table 3 provides a simplified dimension description (good, adequate, or poor) for each website based on the evaluation measure scores, defined differently for each dimension: quality where good $\geq$ 4, adequate=3 to 4, and poor $\leq$ 3 (mean 1-5 rating scale); usability where good $\geq$ 80, adequate=70 to 80, and poor $\leq$ 70 (percentage understandable or actionable); readability, where good $\leq$ 10, adequate=10 to 12, and poor $\geq$ 12 (grade levels; Table 3). The rationale behind the cutoffs for these quality dimensions was based on how difficult it was for the websites to attain the recommended levels for each measure. For website quality, *Good* was used to describe websites that received a rating of  $>4$  on the DISCERN measure as most websites were not able to achieve this. Similarly, with readability, most websites were not able to meet the recommended reading level (grade 7-8); thus, they were rated *Good* if they achieved a reading level of  $<10$ .

**Table 3.** Website dimension comparison<sup>a</sup>.

Website (country)	Search engine order	Overall quality <sup>b</sup>	Usability <sup>c</sup>		Readability <sup>d</sup>
			Understandability score	Actionability score	
HelpGuide (United States)	1	Adequate	Adequate	Poor	Good
National Institute on Aging (United States)	2	Adequate	Poor	Poor	Adequate
National Institute of Mental Health (United States)	3	Adequate	Adequate	Poor	Adequate
Beyond Blue (Australia)	4	Good	Poor	Poor	Poor
Mental Health America (United States)	5	Adequate	Poor	Good	Adequate
WebMD (United States)	6	Adequate	Poor	Poor	Good
Royal College of Psychiatrists (United Kingdom)	8	Good	Poor	Poor	Good
MedlinePlus (United States)	9	Poor	Poor	Poor	Good
Centers for Disease Control and Prevention (United States)	10	Poor	Poor	Poor	Adequate
Healthline (United States)	11	Adequate	Adequate	Poor	Adequate
American Psychological Association (United States)	12	Poor	Poor	Poor	Adequate
Johns Hopkins Medicine (United States)	13	Poor	Poor	Poor	Adequate
Black Dog Institute (Australia)	16	Poor	Adequate	Poor	Poor
Health in Aging (United States)	17	Good	Poor	Poor	Adequate
Age UK (United Kingdom)	24	Poor	Poor	Poor	Good
World Health Organization	25	Poor	Poor	Poor	Poor
Canadian Coalition for Senior's Mental Health (Canada)	28	Adequate	Good	Poor	Good
Substance Abuse and Mental Health Services (United States)	29	Adequate	Poor	Poor	Poor
HealthLinkBC (Canada)	30	Adequate	Adequate	Good	Good

<sup>a</sup>Each website was rated on each dimension as good, adequate, or poor, defined differently for each dimension.

<sup>b</sup>Overall quality was measured by the DISCERN, where good  $\geq$ 4, adequate=3 to 4, and poor  $\leq$ 3 (mean 1-5 rating scale).

<sup>c</sup>Usability was measured by the Patient Education Materials Assessment Tool, where good  $\geq$ 80, adequate=70 to 80, and poor  $\leq$ 70 (percentage understandable or actionable).

<sup>d</sup>Readability was measured by the Simple Measure of Gobbledygook, where good $\geq$ 10, adequate=10 to 12, and poor  $\leq$ 12 (grade levels).

## Discussion

### Principal Findings and Comparison With Prior Work

The purpose of this study was to evaluate the overall quality of LLD information provided by websites, as evaluated by standardized measures of website quality (DISCERN), usability (PEMAT), and readability (SMOG). A secondary purpose of this study was to examine the relationships among website quality, usability, and readability. Website quality ranged from low to high when examining the DISCERN mean scores. Furthermore, when looking at the average of the overall DISCERN rating (item 16), websites were of moderate quality (3.2/5), which indicates that LLD website quality is poor to adequate, providing support for the hypothesis that website quality would be poor to adequate. This finding is consistent with previous website evaluation studies [48,57-59,61].

There were several high-quality websites, with *the Royal College of Psychiatrists* being identified as the website presenting the highest quality information according to DISCERN (4.8). This website also obtained a recommended reading level (grade 7.1), although usability (69.2%) and actionability (50%) were poor. The *Royal College of Psychiatrists* clearly identified the aims of the website and the sources used to create the information according to DISCERN. The same cannot be said for general trends across websites as the mean DISCERN score indicating identification of sources (item 4; “Is it clear what sources of information were used to compile the publication?”) was one of the lowest scores. Sources or references used to create web-based information should be clearly identified as consumers perceive them as positive content indicators, highlighting the trustworthiness and transparency of web-based health information [81].

The *Royal College of Psychiatrists* provided relevant information on comorbid problems that occur late in life and how they might interact with depression, such as physical symptoms, long-term health problems, cognitive issues, and loneliness. This trend was also seen across other evaluated websites more generally, as they provided content that was highly relevant to depression in older adults (eg, sections addressing dementia, vascular depression, and insomnia), which was supported by one of the highest mean DISCERN scores evaluating relevance (item 3; “Is it relevant?”). Providing age-relevant depression information is particularly important as it can aid in increased knowledge about symptom recognition and differentiation of presenting problems (depression vs dementia), which has been shown to lead to more timely access to mental health services [82,83]. This finding is especially relevant for older adults in light of research demonstrating their lower levels of mental health literacy [84,85].

Furthermore, the *Royal College of Psychiatrists* provided a range of unbiased treatment options, such as talk, medication, and complementary and brain stimulation treatments. This trend was also observed across all websites according to the high mean DISCERN score, indicating that the content was unbiased (item 6; “Is it balanced and unbiased?”). The *Royal College of Psychiatrists* included specific information to encourage the involvement of caregivers, specifically how to identify

depression, encourage help seeking, and improve communication between caregivers and recipients. More broadly, information to encourage shared decision-making (eg, sections about talking with physicians or how to help someone with depression) had one of the highest scores across websites according to the mean DISCERN score evaluating support for shared decision-making (item 15; “Does it provide support for shared decision-making?”).

This finding is important to note, as prior research supports older adults’ preference for involvement in decisions related to their health and treatment options [86], and caregivers have been identified as an important part of this process [36]. The inclusion of this information on websites could serve to encourage and improve the shared decision-making processes among older adults, caregivers, and health care providers [87]. As shown previously, websites that provide evidence-based information, diverse treatment options, and content that supports shared decision-making constitute a high-quality website [36,37,52].

Websites generally failed to provide high-quality information about how different treatments work and the risks and benefits associated with those treatment options. A key premise of the shared decision-making process is to provide patients with different treatment options and allow them to weigh the risks and benefits associated with those treatments [16]. Incomplete treatment information may bias the shared decision-making process and limit individuals’ ability to fully weigh their treatment options [88]. Treatment information that clearly defines how treatments work and the associated risks and benefits serve to enhance mental health literacy, consumer empowerment, and shared decision-making [88-90]. Furthermore, when consumers have timely access to balanced, evidence-based information, they can make treatment decisions that align with their preferences [91].

Most websites did not meet the minimum recommended levels of understandability and actionability. Websites with higher usability used formatting that facilitated better understanding, such as bolded main headings and subheadings, large fonts or the option to increase the font, information in short paragraphs or bullet points, boxes to highlight important information, and presenting most information on a single page. Despite providing formatting that promoted website usability, some websites had definitions of medical terms as hyperlinks, forcing individuals to navigate away from the original page the search brought them to. It was also difficult to locate treatment sections on certain websites (eg, *Beyond Blue* and *Mental Health America*), as they were not clearly marked or listed under the adult depression sections. These findings are an especially meaningful aspect of website quality for older adult populations as they may have different needs when using websites, such as the need for increased font size, darker letters with lighter backgrounds, short sections of information, and limited navigational steps when searching [92].

Websites identified actions that individuals could perform to engage in their depression treatment (eg, behavioral strategies); however, most did not go on to further explain the steps and how to complete the recommended actions, limiting consumers’

ability to take the next steps in the help-seeking process. Although the understandability section of the PEMAT was not significantly correlated with DISCERN, the actionability section was significantly positively correlated with DISCERN. Specifically, as the quality of information increased, the actionability of the provided information increased. This finding suggests that higher-quality websites included information that encourages individuals to take action regarding their treatment, ideally resulting in prompt access to mental health services.

Most websites did not meet the recommended reading level (grade 7-8), and therefore, the readability of the websites was poor overall, providing support for the hypothesis that the reading level will be higher than the recommended levels for health information (grades 7-8). This finding is in line with other web-based mental health information evaluation studies that have also found low readability levels among websites [36,48,50,54,60,93]. Website quality was not found to be associated with website readability. A possible explanation for this is that the websites used more complex medical terms to provide higher-quality information but at the cost of an increased reading level. This relationship has been observed in a recent study evaluating the reliability, readability, and quality of hip impingement information on the internet [94]. Furthermore, it is of interest that most of the websites failed to define complex medical terms. It is important to use simple and clear written content and provide definitions for more complex terms to make the content accessible to individuals at all reading levels.

### Limitations

Although this study addresses significant gaps in website evaluation research, there are a number of limitations to note. First, search terms may not be representative of all web-based search strategies used to access information on LLD. Second, the websites included in this study were found by a routine Google search and do not represent the entirety of the websites that could be available to provide information on LLD. Third, raters' pre-existing knowledge about the quality of particular organizations' websites may have also biased raters' perceptions of the quality and usability of the measures used in this study.

### Practice Implications and Recommendations

This study identifies high-quality websites and provides valuable insights into which websites older adults and their caregivers should access to receive high-quality information about treatment options. It also highlights websites providers can recommend to their clients as a resource, such as the *Royal College of Psychiatrists*, *Beyond Blue*, and *Health in Aging*.

Consumers of LLD information should seek web-based resources that discuss the impact of comorbid health problems and associated treatment options specific to this population. Consumers are also encouraged to engage with websites that provide clear evidence-based sources, as the identified sources can be an indicator of transparency and trustworthiness. Older adults and caregivers should seek websites that include content encouraging engagement in shared decision-making, such as how to discuss treatment options with their physician or involve a caregiver in their treatment.

Finally, these findings provide guidance for organizations and website developers to consider when designing a website for older adults. Developers should consider using bolded headings, larger fonts, bullet points, short paragraphs, and text boxes to increase understandability and comprehension of the content. Moreover, developers should consider making navigation between adult depression pages and pages specific to older adults more cohesive, as well as sections of importance (eg, treatment sections) easy to find. Regarding content, developers should further break down actions into small measurable steps encouraging consumers to engage in behavioral changes or treatment seeking and provide tools to aid individuals in taking action (eg, treatment planning sheet). They should use simple and understandable language and provide in-text definitions or glossaries for more complex terms, subsequently increasing information accessibility to individuals with diverse educational backgrounds.

### Future Research

As more people use the internet to access information about mental health problems, it is imperative to understand the quality of websites on the internet to understand whether older adults, caregivers, and health care providers are accessing easy-to-use, accurate, and comprehensive resources. Future research should examine older adults' search strategies and, more specifically, whether older adults identify themselves as "older" when searching for mental health information on the web. Regardless of how older adults identify themselves, websites should better structure how their information is categorized on their websites to ensure that older adults are accessing the information relevant to them.

It will also be important to look more in depth at the usability characteristics of websites that the PEMAT did not address. A measure such as the Visual Aesthetics of Websites Inventory, which examines more specific aesthetics of websites such as simplicity, diversity, colorfulness, and craftsmanship, would be useful in further evaluating usability [95]. Future researchers should fully evaluate the aesthetics of websites as it has implications on individuals' first impressions and whether they will revisit the site or recommend it to others [96].

### Conclusions

This is the first study to examine the quality of LLD information on the internet, and it addresses a gap in the literature by highlighting the quality of several LLD websites accessible on the internet. This study took a multifaceted approach to measuring website quality by using multiple measures to better understand different aspects that contribute to the overall quality of websites. The quality of LLD websites varied, ranging from low to high quality. Overall, the quality of the websites was adequate, and the usability and reading levels of the websites were poor. Websites provided information about particular problems that may affect depression in later life but lacked key information on how treatments work and the risks and benefits associated with treatments. Treatment sections were difficult to navigate or were found under adult depression sections. The ability of the websites to encourage understanding and action in individuals was also poor, and the information presented was higher than the recommended reading level. Websites were

strong at providing multiple treatment options relevant to older adults and encouraging shared decision-making. They provided visual cues and formatting, which facilitated better understanding (eg, use of bolded headings, short paragraphs, or bullet points), and some websites were able to attain the recommended reading level. Website developers should consider increasing the quality, usability, and readability to produce high-quality information for older adults. High-quality websites may increase the mental health literacy of older adults and

caregivers and improve the shared decision-making process. Health care providers should be aware of high-quality websites and should incorporate the use of high-quality websites into the shared decision-making process. They should direct older adults and caregivers to the high-quality websites identified in this study and use them as a decision-making tool by directing them to sections presenting different treatment options to further discuss in their ongoing care.

---

## Acknowledgments

The authors would like to acknowledge funding from the University of Manitoba Centre on Aging (KAR, principal investigator) and Research Manitoba (KAR, principal investigator) for the completion and publication of this study.

---

## Authors' Contributions

TAMP completed data collection, analyses, and writing. KAR participated in conceptualization, methodology, investigation, resources, supervision, and writing and editing of the manuscript. PLK participated in data collection and analyses. MTB participated in the conceptualization, methodology, supervision, and editing of the manuscript. All authors reviewed and approved the final manuscript.

---

## Conflicts of Interest

None declared.

---

## References

1. Alexopoulos GS. Depression in the elderly. *Lancet* 2005;365(9475):1961-1970. [doi: [10.1016/S0140-6736\(05\)66665-2](https://doi.org/10.1016/S0140-6736(05)66665-2)] [Medline: [15936426](https://pubmed.ncbi.nlm.nih.gov/15936426/)]
2. American Psychiatric Association. *Diagnostic And Statistical Manual Of Mental Disorders* 5th edition. Virginia, United States: American Psychiatric Association; 2013.
3. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *Int J Methods Psychiatr Res* 2012 Sep;21(3):169-184 [FREE Full text] [doi: [10.1002/mpr.1359](https://doi.org/10.1002/mpr.1359)] [Medline: [22865617](https://pubmed.ncbi.nlm.nih.gov/22865617/)]
4. Pearson C, Janz T, Ali J. Mental and substance use disorders in Canada. *Statistics Canada*. URL: <http://www.statcan.gc.ca/pub/82-624-x/2013001/article/11855-eng.htm#n9> [accessed 2022-09-02]
5. Reynolds K, Pietrzak RH, El-Gabalawy R, Mackenzie CS, Sareen J. Prevalence of psychiatric disorders in U.S. older adults: findings from a nationally representative survey. *World Psychiatry* 2015 Mar;14(1):74-81 [FREE Full text] [doi: [10.1002/wps.20193](https://doi.org/10.1002/wps.20193)] [Medline: [25655161](https://pubmed.ncbi.nlm.nih.gov/25655161/)]
6. Gum AM, King-Kallimanis B, Kohn R. Prevalence of mood, anxiety, and substance-abuse disorders for older Americans in the national comorbidity survey-replication. *Am J Geriatr Psychiatry* 2009 Oct;17(9):769-781. [doi: [10.1097/JGP.0b013e3181ad4f5a](https://doi.org/10.1097/JGP.0b013e3181ad4f5a)] [Medline: [19700949](https://pubmed.ncbi.nlm.nih.gov/19700949/)]
7. Laborde-Lahoz P, El-Gabalawy R, Kinley J, Kirwin PD, Sareen J, Pietrzak RH. Subsyndromal depression among older adults in the USA: prevalence, comorbidity, and risk for new-onset psychiatric disorders in late life. *Int J Geriatr Psychiatry* 2015 Jul;30(7):677-685. [doi: [10.1002/gps.4204](https://doi.org/10.1002/gps.4204)] [Medline: [25345806](https://pubmed.ncbi.nlm.nih.gov/25345806/)]
8. *Preventing Suicide: A Global Imperative*. Geneva: World Health Organization; 2014.
9. Fiske A, Wetherell JL, Gatz M. Depression in older adults. *Annu Rev Clin Psychol* 2009;5:363-389 [FREE Full text] [doi: [10.1146/annurev.clinpsy.032408.153621](https://doi.org/10.1146/annurev.clinpsy.032408.153621)] [Medline: [19327033](https://pubmed.ncbi.nlm.nih.gov/19327033/)]
10. Farrer L, Leach L, Griffiths KM, Christensen H, Jorm AF. Age differences in mental health literacy. *BMC Public Health* 2008 May 20;8:125. [doi: [10.1186/1471-2458-8-125](https://doi.org/10.1186/1471-2458-8-125)] [Medline: [18423049](https://pubmed.ncbi.nlm.nih.gov/18423049/)]
11. Wetherell JL, Petkus AJ, McChesney K, Stein MB, Judd PH, Rockwell E, et al. Older adults are less accurate than younger adults at identifying symptoms of anxiety and depression. *J Nerv Ment Dis* 2009 Aug;197(8):623-626 [FREE Full text] [doi: [10.1097/NMD.0b013e3181b0c081](https://doi.org/10.1097/NMD.0b013e3181b0c081)] [Medline: [19684501](https://pubmed.ncbi.nlm.nih.gov/19684501/)]
12. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. "Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. *Med J Aust* 1997 Mar 17;166(4):182-186. [doi: [10.5694/j.1326-5377.1997.tb140071.x](https://doi.org/10.5694/j.1326-5377.1997.tb140071.x)] [Medline: [9066546](https://pubmed.ncbi.nlm.nih.gov/9066546/)]
13. Kim YS, Lee HY, Lee MH, Simms T, Park BH. Mental health literacy in Korean older adults: a cross-sectional survey. *J Psychiatr Ment Health Nurs* 2017 Oct;24(7):523-533. [doi: [10.1111/jpm.12395](https://doi.org/10.1111/jpm.12395)] [Medline: [28474412](https://pubmed.ncbi.nlm.nih.gov/28474412/)]

14. Mackenzie CS, Scott T, Mather A, Sareen J. Older adults' help-seeking attitudes and treatment beliefs concerning mental health problems. *Am J Geriatr Psychiatry* 2008 Dec;16(12):1010-1019 [FREE Full text] [doi: [10.1097/JGP.0b013e31818cd3be](https://doi.org/10.1097/JGP.0b013e31818cd3be)] [Medline: [19038900](https://pubmed.ncbi.nlm.nih.gov/19038900/)]
15. Xie B, Wang M, Feldman R, Zhou L. Exploring older and younger adults' preferences for health information and participation in decision making using the Health Information Wants Questionnaire (HIWQ). *Health Expect* 2014 Dec;17(6):795-808 [FREE Full text] [doi: [10.1111/j.1369-7625.2012.00804.x](https://doi.org/10.1111/j.1369-7625.2012.00804.x)] [Medline: [22783800](https://pubmed.ncbi.nlm.nih.gov/22783800/)]
16. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). *Soc Sci Med* 1997 Mar;44(5):681-692. [doi: [10.1016/s0277-9536\(96\)00221-3](https://doi.org/10.1016/s0277-9536(96)00221-3)] [Medline: [9032835](https://pubmed.ncbi.nlm.nih.gov/9032835/)]
17. Fisher LJ, Goldney RD. Differences in community mental health literacy in older and younger Australians. *Int J Geriatr Psychiatry* 2003 Jan;18(1):33-40. [doi: [10.1002/gps.769](https://doi.org/10.1002/gps.769)] [Medline: [12497554](https://pubmed.ncbi.nlm.nih.gov/12497554/)]
18. Loh A, Simon D, Wills CE, Kriston L, Niebling W, Härter M. The effects of a shared decision-making intervention in primary care of depression: a cluster-randomized controlled trial. *Patient Educ Couns* 2007 Aug;67(3):324-332. [doi: [10.1016/j.pec.2007.03.023](https://doi.org/10.1016/j.pec.2007.03.023)] [Medline: [17509808](https://pubmed.ncbi.nlm.nih.gov/17509808/)]
19. Siegel CA. Embracing the internet for progress in shared decision-making. *Inflamm Bowel Dis* 2007 Dec;13(12):1579-1580. [doi: [10.1002/ibd.20259](https://doi.org/10.1002/ibd.20259)] [Medline: [17879279](https://pubmed.ncbi.nlm.nih.gov/17879279/)]
20. 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](https://doi.org/10.1371/journal.pone.0234663)] [Medline: [32530937](https://pubmed.ncbi.nlm.nih.gov/32530937/)]
21. Beck F, Richard J, Nguyen-Thanh V, Montagni I, Parizot I, Renahy E. Use of the internet as a health information resource among French young adults: results from a nationally representative survey. *J Med Internet Res* 2014 May 13;16(5):e128 [FREE Full text] [doi: [10.2196/jmir.2934](https://doi.org/10.2196/jmir.2934)] [Medline: [24824164](https://pubmed.ncbi.nlm.nih.gov/24824164/)]
22. Ducrot P, Montagni I, Nguyen Thanh V, Serry A, Richard J. Evolution of online health-related information seeking in France from 2010 to 2017: results from nationally representative surveys. *J Med Internet Res* 2021 Apr 14;23(4):e18799 [FREE Full text] [doi: [10.2196/18799](https://doi.org/10.2196/18799)] [Medline: [33851927](https://pubmed.ncbi.nlm.nih.gov/33851927/)]
23. Shahab L, Brown J, Gardner B, Smith SG. Seeking health information and support online: does it differ as a function of engagement in risky health behaviors? Evidence from the health information national trends survey. *J Med Internet Res* 2014 Dec 06;16(11):e253 [FREE Full text] [doi: [10.2196/jmir.3368](https://doi.org/10.2196/jmir.3368)] [Medline: [25380308](https://pubmed.ncbi.nlm.nih.gov/25380308/)]
24. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. *J Med Internet Res* 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: [10.2196/11224](https://doi.org/10.2196/11224)] [Medline: [30882361](https://pubmed.ncbi.nlm.nih.gov/30882361/)]
25. Getting a second opinion: health information and the internet. Statistics Canada. 2015. URL: <https://www150.statcan.gc.ca/n1/pub/82-003-x/2008001/article/10515/5002590-eng.htm> [accessed 2022-09-02]
26. Brunault P, Bray A, Rerolle C, Cognet S, Gaillard P, El-Hage W. [Differences in access to Internet and Internet-based information seeking according to the type of psychiatric disorder]. *Rev Epidemiol Sante Publique* 2017 May;65(2):125-136. [doi: [10.1016/j.respe.2016.11.001](https://doi.org/10.1016/j.respe.2016.11.001)] [Medline: [28153646](https://pubmed.ncbi.nlm.nih.gov/28153646/)]
27. Kalkreuth S, Trefflich F, Rummel-Kluge C. Mental health related Internet use among psychiatric patients: a cross-sectional analysis. *BMC Psychiatry* 2014 Dec 24;14:368 [FREE Full text] [doi: [10.1186/s12888-014-0368-7](https://doi.org/10.1186/s12888-014-0368-7)] [Medline: [25599722](https://pubmed.ncbi.nlm.nih.gov/25599722/)]
28. Scott J, Hockey S, Ospina-Pinillos L, Doraiswamy PM, Alvarez-Jimenez M, Hickie I. Research to Clinical Practice-Youth seeking mental health information online and its impact on the first steps in the patient journey. *Acta Psychiatr Scand* 2022 Mar;145(3):301-314. [doi: [10.1111/acps.13390](https://doi.org/10.1111/acps.13390)] [Medline: [34923619](https://pubmed.ncbi.nlm.nih.gov/34923619/)]
29. Villani M, Kovess-Masfety V. How do people experiencing schizophrenia spectrum disorders or other psychotic disorders use the internet to get information on their mental health? Literature review and recommendations. *JMIR Ment Health* 2017 Jan 03;4(1):e1 [FREE Full text] [doi: [10.2196/mental.5946](https://doi.org/10.2196/mental.5946)] [Medline: [28049620](https://pubmed.ncbi.nlm.nih.gov/28049620/)]
30. Nikoloudakis IA, Vandelanotte C, Rebar AL, Schoeppe S, Alley S, Duncan MJ, et al. Examining the correlates of online health information-seeking behavior among men compared with women. *Am J Mens Health* 2018 Sep;12(5):1358-1367 [FREE Full text] [doi: [10.1177/1557988316650625](https://doi.org/10.1177/1557988316650625)] [Medline: [27193765](https://pubmed.ncbi.nlm.nih.gov/27193765/)]
31. Internet/broadband fact sheet. Pew Research Center. 2021 Apr 7. URL: <https://www.pewresearch.org/internet/fact-sheet/internet-broadband/> [accessed 2022-08-31]
32. Yoon H, Jang Y, Vaughan PW, Garcia M. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. *J Appl Gerontol* 2020 Jan;39(1):105-110. [doi: [10.1177/0733464818770772](https://doi.org/10.1177/0733464818770772)] [Medline: [29661052](https://pubmed.ncbi.nlm.nih.gov/29661052/)]
33. Crabb RM, Rafie S, Weingardt KR. Health-related internet use in older primary care patients. *Gerontology* 2012;58(2):164-170. [doi: [10.1159/000329340](https://doi.org/10.1159/000329340)] [Medline: [21734360](https://pubmed.ncbi.nlm.nih.gov/21734360/)]
34. Bauer R, Glenn T, Strejilevich S, Conell J, Alda M, Ardaur R, Miranda-Scippa, et al. Internet use by older adults with bipolar disorder: international survey results. *Int J Bipolar Disord* 2018 Sep 04;6(1):20 [FREE Full text] [doi: [10.1186/s40345-018-0127-7](https://doi.org/10.1186/s40345-018-0127-7)] [Medline: [30178112](https://pubmed.ncbi.nlm.nih.gov/30178112/)]
35. Moulton A, Burroughs H, Kingstone T, Chew-Graham CA. How older adults self-manage distress - does the internet have a role? A qualitative study. *BMC Fam Pract* 2018 Nov 29;19(1):185 [FREE Full text] [doi: [10.1186/s12875-018-0874-7](https://doi.org/10.1186/s12875-018-0874-7)] [Medline: [30497414](https://pubmed.ncbi.nlm.nih.gov/30497414/)]
36. Elliott J, McNeil H, Ashbourne J, Huson K, Boscart V, Stolee P. Engaging older adults in health care decision-making: a realist synthesis. *Patient* 2016 Oct;9(5):383-393 [FREE Full text] [doi: [10.1007/s40271-016-0168-x](https://doi.org/10.1007/s40271-016-0168-x)] [Medline: [27048393](https://pubmed.ncbi.nlm.nih.gov/27048393/)]

37. Funk LM, Dansereau L, Novek S. Carers as system navigators: exploring sources, processes and outcomes of structural burden. *Gerontologist* 2019 May 17;59(3):426-435. [doi: [10.1093/geront/gnx175](https://doi.org/10.1093/geront/gnx175)] [Medline: [29165584](https://pubmed.ncbi.nlm.nih.gov/29165584/)]
38. Bangerter LR, Griffin J, Harden K, Rutten LJ. Health information-seeking behaviors of family caregivers: analysis of the health information national trends survey. *JMIR Aging* 2019 Jan 14;2(1):e11237 [FREE Full text] [doi: [10.2196/11237](https://doi.org/10.2196/11237)] [Medline: [31518309](https://pubmed.ncbi.nlm.nih.gov/31518309/)]
39. Fox L, Brenner J. Family caregivers online. Pew Research Center. 2012 Jul 12. URL: <https://www.pewresearch.org/internet/2012/07/12/family-caregivers-online/> [accessed 2022-04-10]
40. Shaffer KM, Chow PI, Cohn WF, Ingersoll KS, Ritterband LM. Informal caregivers' use of internet-based health resources: an analysis of the health information national trends survey. *JMIR Aging* 2018 Dec 18;1(2):e11051 [FREE Full text] [doi: [10.2196/11051](https://doi.org/10.2196/11051)] [Medline: [31518244](https://pubmed.ncbi.nlm.nih.gov/31518244/)]
41. Cline RJ, Haynes KM. Consumer health information seeking on the internet: the state of the art. *Health Educ Res* 2001 Dec;16(6):671-692. [doi: [10.1093/her/16.6.671](https://doi.org/10.1093/her/16.6.671)] [Medline: [11780707](https://pubmed.ncbi.nlm.nih.gov/11780707/)]
42. Khazaal Y, Chatton A, Cochand S, Hoch A, Khankarli MB, Khan R, et al. Internet use by patients with psychiatric disorders in search for general and medical informations. *Psychiatr Q* 2008 Dec;79(4):301-309. [doi: [10.1007/s11126-008-9083-1](https://doi.org/10.1007/s11126-008-9083-1)] [Medline: [18756354](https://pubmed.ncbi.nlm.nih.gov/18756354/)]
43. Ek S, Eriksson-Backa K, Niemelä R. Use of and trust in health information on the internet: a nationwide eight-year follow-up survey. *Inform Health Soc Care* 2013 Sep;38(3):236-245. [doi: [10.3109/17538157.2013.764305](https://doi.org/10.3109/17538157.2013.764305)] [Medline: [23514041](https://pubmed.ncbi.nlm.nih.gov/23514041/)]
44. Fiksdal AS, Kumbamu A, Jadhav AS, Cocos C, Nelsen LA, Pathak J, et al. Evaluating the process of online health information searching: a qualitative approach to exploring consumer perspectives. *J Med Internet Res* 2014 Oct 07;16(10):e224 [FREE Full text] [doi: [10.2196/jmir.3341](https://doi.org/10.2196/jmir.3341)] [Medline: [25348028](https://pubmed.ncbi.nlm.nih.gov/25348028/)]
45. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, et al. Older adults talk technology: technology usage and attitudes. *Comput Human Behav* 2010 Nov 01;26(6):1710-1721 [FREE Full text] [doi: [10.1016/j.chb.2010.06.020](https://doi.org/10.1016/j.chb.2010.06.020)] [Medline: [20967133](https://pubmed.ncbi.nlm.nih.gov/20967133/)]
46. Perrin A, Atske S. About three-in-ten U.S. adults say they are 'almost constantly' online. Pew Research Center. 2021 Mar 26. URL: <https://www.pewresearch.org/fact-tank/2021/03/26/about-three-in-ten-u-s-adults-say-they-are-almost-constantly-online/> [accessed 2022-09-02]
47. Murray E, Lo B, Pollack L, Donelan K, Catania J, Lee K, et al. The impact of health information on the internet on health care and the physician-patient relationship: national U.S. survey among 1,050 U.S. physicians. *J Med Internet Res* 2003;5(3):e17 [FREE Full text] [doi: [10.2196/jmir.5.3.e17](https://doi.org/10.2196/jmir.5.3.e17)] [Medline: [14517108](https://pubmed.ncbi.nlm.nih.gov/14517108/)]
48. Reavley NJ, Jorm AF. The quality of mental disorder information websites: a review. *Patient Educ Couns* 2011 Nov;85(2):e16-e25. [doi: [10.1016/j.pec.2010.10.015](https://doi.org/10.1016/j.pec.2010.10.015)] [Medline: [21087837](https://pubmed.ncbi.nlm.nih.gov/21087837/)]
49. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA* 2002;287(20):2691-2700. [doi: [10.1001/jama.287.20.2691](https://doi.org/10.1001/jama.287.20.2691)] [Medline: [12020305](https://pubmed.ncbi.nlm.nih.gov/12020305/)]
50. Kim H, Xie B. Health literacy in the eHealth era: a systematic review of the literature. *Patient Educ Couns* 2017 Jun;100(6):1073-1082. [doi: [10.1016/j.pec.2017.01.015](https://doi.org/10.1016/j.pec.2017.01.015)] [Medline: [28174067](https://pubmed.ncbi.nlm.nih.gov/28174067/)]
51. Fu K, Wong PW, Yip PS. What do internet users seek to know about depression from web searches? A descriptive study of 21 million web queries. *J Clin Psychiatry* 2010 Sep;71(9):1246-1247. [doi: [10.4088/JCP.09105903blu](https://doi.org/10.4088/JCP.09105903blu)] [Medline: [20923627](https://pubmed.ncbi.nlm.nih.gov/20923627/)]
52. Reichow B, Halpern JI, Steinhoff TB, Letsinger N, Naples A, Volkmar FR. Characteristics and quality of autism websites. *J Autism Dev Disord* 2012 Jun;42(6):1263-1274. [doi: [10.1007/s10803-011-1342-6](https://doi.org/10.1007/s10803-011-1342-6)] [Medline: [21822761](https://pubmed.ncbi.nlm.nih.gov/21822761/)]
53. Tsakonas G, Papatheodorou C. Analysing and evaluating usefulness and usability in electronic information services. *J Inform Sci* 2006 Jun 09;32(5):400-419. [doi: [10.1177/0165551506065934](https://doi.org/10.1177/0165551506065934)]
54. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the internet: accessibility, quality, and readability in English and Spanish. *JAMA* 2001;285(20):2612-2621 [FREE Full text] [doi: [10.1001/jama.285.20.2612](https://doi.org/10.1001/jama.285.20.2612)] [Medline: [11368735](https://pubmed.ncbi.nlm.nih.gov/11368735/)]
55. Coquard O, Fernandez S, Zullino D, Khazaal Y. A follow-up study on the quality of alcohol dependence-related information on the web. *Subst Abuse Treat Prev Policy* 2011 Jun 10;6:13 [FREE Full text] [doi: [10.1186/1747-597X-6-13](https://doi.org/10.1186/1747-597X-6-13)] [Medline: [21663650](https://pubmed.ncbi.nlm.nih.gov/21663650/)]
56. Reavley NJ, Mackinnon AJ, Morgan AJ, Alvarez-Jimenez M, Hetrick SE, Killackey E, et al. Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. *Psychol Med* 2012 Aug;42(8):1753-1762. [doi: [10.1017/S003329171100287X](https://doi.org/10.1017/S003329171100287X)] [Medline: [22166182](https://pubmed.ncbi.nlm.nih.gov/22166182/)]
57. Ferreira-Lay P, Miller S. The quality of internet information on depression for lay people. *Psychiatr Bull* 2018 Jan 02;32(5):170-173. [doi: [10.1192/pb.bp.107.016188](https://doi.org/10.1192/pb.bp.107.016188)]
58. Griffiths KM, Christensen H. Quality of web based information on treatment of depression: cross sectional survey. *BMJ* 2000 Dec 16;321(7275):1511-1515 [FREE Full text] [doi: [10.1136/bmj.321.7275.1511](https://doi.org/10.1136/bmj.321.7275.1511)] [Medline: [11118181](https://pubmed.ncbi.nlm.nih.gov/11118181/)]
59. Lissman TL, Boehnlein JK. A critical review of internet information about depression. *Psychiatr Serv* 2001 Aug;52(8):1046-1050. [doi: [10.1176/appi.ps.52.8.1046](https://doi.org/10.1176/appi.ps.52.8.1046)] [Medline: [11474049](https://pubmed.ncbi.nlm.nih.gov/11474049/)]

60. Walsh K, Pryor TA, Reynolds KA, Walker JR, Mobilizing Minds Research Group. Searching for answers: how well do depression websites answer the public's questions about treatment choices? *Patient Educ Couns* 2019 Jan;102(1):99-105. [doi: [10.1016/j.pec.2018.08.019](https://doi.org/10.1016/j.pec.2018.08.019)] [Medline: [30150127](https://pubmed.ncbi.nlm.nih.gov/30150127/)]
61. Zermatten A, Khazaal Y, Coquard O, Chatton A, Bondolfi G. Quality of web-based information on depression. *Depress Anxiety* 2010 Sep;27(9):852-858. [doi: [10.1002/da.20665](https://doi.org/10.1002/da.20665)] [Medline: [20099271](https://pubmed.ncbi.nlm.nih.gov/20099271/)]
62. Charnock D. *The DISCERN Handbook Quality Criteria for Consumer Health Information on Treatment Choices*. Abingdon, Oxon: Radcliffe Medical Press; 1998.
63. Shoemaker SJ, Wolf MS, Brach C. The patient education materials assessment tool (PEMAT) and user's guide. Agency for Healthcare Research and Quality. 2020. URL: <https://www.ahrq.gov/professionals/prevention-chronic-care/improve/self-mgmt/pemat/index.html> [accessed 2022-09-02]
64. McLaughlin G. SMOG grading: a new readability formula. *J Read* 1969;12(8):639-646.
65. Ratcliff C. What are the top 10 most popular search engines? *Search Engine Watch*. 2016 Aug 8. URL: <https://searchenginewatch.com/2016/08/08/what-are-the-top-10-most-popular-search-engines/> [accessed 2022-09-02]
66. Sbaffi L, Zhao C. Modeling the online health information seeking process: information channel selection among university students. *J Assoc Inform Sci Technol* 2019 Apr 13;71(2):196-207 [FREE Full text] [doi: [10.1002/asi.24230](https://doi.org/10.1002/asi.24230)]
67. iProspect blended search results study. *iProspect*. 2008 Apr. URL: <https://tinyurl.com/4m9rb58b> [accessed 2022-08-31]
68. Barnes C, Harvey R, Wilde A, Hadzi-Pavlovic D, Wilhelm K, Mitchell PB. Review of the quality of information on bipolar disorder on the internet. *Aust N Z J Psychiatry* 2009 Jan 01;43(10):934-945. [doi: [10.1080/00048670903179137](https://doi.org/10.1080/00048670903179137)]
69. Batchelor JM, Ohya Y. Use of the DISCERN instrument by patients and health professionals to assess information resources on treatments for asthma and atopic dermatitis. *Allergol Int* 2009 Mar;58(1):141-145 [FREE Full text] [doi: [10.2332/allergolint.08-SC-0022](https://doi.org/10.2332/allergolint.08-SC-0022)] [Medline: [19153541](https://pubmed.ncbi.nlm.nih.gov/19153541/)]
70. Cerminara C, Santarone ME, Casarelli L, Curatolo P, El Malhany N. Use of the DISCERN tool for evaluating web searches in childhood epilepsy. *Epilepsy Behav* 2014 Dec;41:119-121. [doi: [10.1016/j.yebeh.2014.09.053](https://doi.org/10.1016/j.yebeh.2014.09.053)] [Medline: [25461201](https://pubmed.ncbi.nlm.nih.gov/25461201/)]
71. Grant N, Rodger S, Hoffmann T. Evaluation of autism-related health information on the web. *J Appl Res Intellect Disabil* 2015 Jul;28(4):276-282. [doi: [10.1111/jar.12127](https://doi.org/10.1111/jar.12127)] [Medline: [25224690](https://pubmed.ncbi.nlm.nih.gov/25224690/)]
72. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 1999 Mar;53(2):105-111 [FREE Full text] [doi: [10.1136/jech.53.2.105](https://doi.org/10.1136/jech.53.2.105)] [Medline: [10396471](https://pubmed.ncbi.nlm.nih.gov/10396471/)]
73. McCool ME, Wahl J, Schlecht I, Apfelfbacher C. Evaluating written patient information for eczema in German: comparing the reliability of two instruments, DISCERN and EQIP. *PLoS One* 2015;10(10):e0139895 [FREE Full text] [doi: [10.1371/journal.pone.0139895](https://doi.org/10.1371/journal.pone.0139895)] [Medline: [26440612](https://pubmed.ncbi.nlm.nih.gov/26440612/)]
74. Rees CE, Ford JE, Sheard CE. Evaluating the reliability of DISCERN: a tool for assessing the quality of written patient information on treatment choices. *Patient Educ Couns* 2002 Jul;47(3):273-275. [doi: [10.1016/s0738-3991\(01\)00225-7](https://doi.org/10.1016/s0738-3991(01)00225-7)] [Medline: [12088606](https://pubmed.ncbi.nlm.nih.gov/12088606/)]
75. Balakrishnan V, Chandy Z, Verma SP. Are online Zenker's diverticulum materials readable and understandable? *Otolaryngol Head Neck Surg* 2016 Nov;155(5):758-763. [doi: [10.1177/0194599816655302](https://doi.org/10.1177/0194599816655302)] [Medline: [27329416](https://pubmed.ncbi.nlm.nih.gov/27329416/)]
76. Morony S, McCaffery KJ, Kirkendall S, Jansen J, Webster AC. Health literacy demand of printed lifestyle patient information materials aimed at people with chronic kidney disease: are materials easy to understand and act on and do they use meaningful visual aids? *J Health Commun* 2017 Feb 25;22(2):163-170. [doi: [10.1080/10810730.2016.1258744](https://doi.org/10.1080/10810730.2016.1258744)] [Medline: [28121226](https://pubmed.ncbi.nlm.nih.gov/28121226/)]
77. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. *Patient Educ Couns* 2014 Sep;96(3):395-403 [FREE Full text] [doi: [10.1016/j.pec.2014.05.027](https://doi.org/10.1016/j.pec.2014.05.027)] [Medline: [24973195](https://pubmed.ncbi.nlm.nih.gov/24973195/)]
78. Vishnevetsky J, Walters CB, Tan KS. Interrater reliability of the patient education materials assessment tool (PEMAT). *Patient Educ Couns* 2018 Mar;101(3):490-496 [FREE Full text] [doi: [10.1016/j.pec.2017.09.003](https://doi.org/10.1016/j.pec.2017.09.003)] [Medline: [28899713](https://pubmed.ncbi.nlm.nih.gov/28899713/)]
79. Wang L, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. *Res Social Adm Pharm* 2013;9(5):503-516. [doi: [10.1016/j.sapharm.2012.05.009](https://doi.org/10.1016/j.sapharm.2012.05.009)] [Medline: [22835706](https://pubmed.ncbi.nlm.nih.gov/22835706/)]
80. Reynolds KA, Walker JR, Walsh K, Mobilizing Minds Research Group. How well do websites concerning children's anxiety answer parents' questions about treatment choices? *Clin Child Psychol Psychiatry* 2015 Oct;20(4):555-569. [doi: [10.1177/1359104514534948](https://doi.org/10.1177/1359104514534948)] [Medline: [24830663](https://pubmed.ncbi.nlm.nih.gov/24830663/)]
81. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. *J Med Internet Res* 2019 May 02;21(5):e12522 [FREE Full text] [doi: [10.2196/12522](https://doi.org/10.2196/12522)] [Medline: [31045507](https://pubmed.ncbi.nlm.nih.gov/31045507/)]
82. Brijnath B, Protheroe J, Mahtani KR, Antoniadis J. Do web-based mental health literacy interventions improve the mental health literacy of adult consumers? Results from a systematic review. *J Med Internet Res* 2016 Jun 20;18(6):e165 [FREE Full text] [doi: [10.2196/jmir.5463](https://doi.org/10.2196/jmir.5463)] [Medline: [27323907](https://pubmed.ncbi.nlm.nih.gov/27323907/)]
83. Kim YS, Rhee TG, Lee HY, Park BH, Sharratt ML. Mental health literacy as a mediator in use of mental health services among older Korean adults. *Int Psychogeriatr* 2016 Oct 20;29(2):269-279. [doi: [10.1017/s1041610216001721](https://doi.org/10.1017/s1041610216001721)]



84. Cronin M, O'Hanlon S, O'Connor M. Readability level of patient information leaflets for older people. *Ir J Med Sci* 2011 Mar;180(1):139-142. [doi: [10.1007/s11845-010-0624-x](https://doi.org/10.1007/s11845-010-0624-x)] [Medline: [21046472](https://pubmed.ncbi.nlm.nih.gov/21046472/)]
85. Nutbeam D, Wise M. Australia: planning for better health. *Promotion Educ* 2016 Aug 24;1(4):19-24. [doi: [10.1177/175797599300100408](https://doi.org/10.1177/175797599300100408)]
86. O'Neal EL, Adams JR, McHugo GJ, Van Citters AD, Drake RE, Bartels SJ. Preferences of older and younger adults with serious mental illness for involvement in decision-making in medical and psychiatric settings. *Am J Geriatr Psychiatry* 2008 Oct;16(10):826-833. [doi: [10.1097/JGP.0b013e318181f992](https://doi.org/10.1097/JGP.0b013e318181f992)] [Medline: [18827229](https://pubmed.ncbi.nlm.nih.gov/18827229/)]
87. Luo A, Qin L, Yuan Y, Yang Z, Liu F, Huang P, et al. The effect of online health information seeking on physician-patient relationships: systematic review. *J Med Internet Res* 2022 Feb 10;24(2):e23354 [FREE Full text] [doi: [10.2196/23354](https://doi.org/10.2196/23354)] [Medline: [35142620](https://pubmed.ncbi.nlm.nih.gov/35142620/)]
88. Thapa DK, Visentin DC, Kornhaber R, West S, Cleary M. The influence of online health information on health decisions: a systematic review. *Patient Educ Couns* 2021 Apr;104(4):770-784. [doi: [10.1016/j.pec.2020.11.016](https://doi.org/10.1016/j.pec.2020.11.016)] [Medline: [33358253](https://pubmed.ncbi.nlm.nih.gov/33358253/)]
89. Alden DL. Decision aid influences on factors associated with patient empowerment prior to cancer treatment decision making. *Med Decis Making* 2014 Oct;34(7):884-898. [doi: [10.1177/0272989X14536780](https://doi.org/10.1177/0272989X14536780)] [Medline: [24903122](https://pubmed.ncbi.nlm.nih.gov/24903122/)]
90. Náfrádi L, Nakamoto K, Csabai M, Papp-Zipernovszky O, Schulz PJ. An empirical test of the Health Empowerment Model: does patient empowerment moderate the effect of health literacy on health status? *Patient Educ Couns* 2018 Mar;101(3):511-517. [doi: [10.1016/j.pec.2017.09.004](https://doi.org/10.1016/j.pec.2017.09.004)] [Medline: [28899712](https://pubmed.ncbi.nlm.nih.gov/28899712/)]
91. Krist AH, Tong ST, Aycock RA, Longo DR. Engaging patients in decision-making and behavior change to promote prevention. *Stud Health Technol Inform* 2017;240:284-302 [FREE Full text] [Medline: [28972524](https://pubmed.ncbi.nlm.nih.gov/28972524/)]
92. Nahm E, Preece J, Resnick B, Mills ME. Usability of health Web sites for older adults: a preliminary study. *Comput Inform Nurs* 2004;22(6):326-34; quiz 335. [doi: [10.1097/00024665-200411000-00007](https://doi.org/10.1097/00024665-200411000-00007)] [Medline: [15602301](https://pubmed.ncbi.nlm.nih.gov/15602301/)]
93. Grohol JM, Slimowicz J, Granda R. The quality of mental health information commonly searched for on the internet. *Cyberpsychol Behav Soc Netw* 2014 Apr;17(4):216-221. [doi: [10.1089/cyber.2013.0258](https://doi.org/10.1089/cyber.2013.0258)] [Medline: [24237287](https://pubmed.ncbi.nlm.nih.gov/24237287/)]
94. Küçükduymaz F, Gomez MM, Secrist E, Parvizi J. Reliability, readability and quality of online information about femoral acetabular impingement. *Arch Bone Jt Surg* 2015 Jul;3(3):163-168 [FREE Full text] [Medline: [26213699](https://pubmed.ncbi.nlm.nih.gov/26213699/)]
95. Moshagen M, Thielsch MT. Facets of visual aesthetics. *Int J Human Comput Stud* 2010 Oct;68(10):689-709. [doi: [10.1016/j.ijhcs.2010.05.006](https://doi.org/10.1016/j.ijhcs.2010.05.006)]
96. Thielsch MT, Blotenberg I, Jaron R. User evaluation of websites: from first impression to recommendation. *Interact Comput* 2013 Jun 04;26(1):89-102. [doi: [10.1093/iwc/iwt033](https://doi.org/10.1093/iwc/iwt033)]

## Abbreviations

**LLD:** late-life depression

**MDD:** major depressive disorder

**PEMAT:** Patient Education Materials Assessment Tool

**SMOG:** Simple Measure of Gobbledygook

*Edited by A Mavragani; submitted 06.01.22; peer-reviewed by A Tuck, S Sutherland; comments to author 15.02.22; revised version received 12.04.22; accepted 04.05.22; published 12.09.22.*

*Please cite as:*

*Pryor TAM, Reynolds KA, Kirby PL, Bernstein MT*

*Quality of Late-Life Depression Information on the Internet: Website Evaluation Study*

*JMIR Form Res* 2022;6(9):e36177

URL: <https://formative.jmir.org/2022/9/e36177>

doi: [10.2196/36177](https://doi.org/10.2196/36177)

PMID: [36094802](https://pubmed.ncbi.nlm.nih.gov/36094802/)

©Teaghan A M Pryor, Kristin A Reynolds, Paige L Kirby, Matthew T Bernstein. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 12.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Home-Based Electronic Cognitive Therapy in Patients With Alzheimer Disease: Feasibility Randomized Controlled Trial

Anna Marin<sup>1,2</sup>, BA; Renée DeCaro<sup>1,2</sup>, PhD; Kylie Schiloski<sup>1</sup>, BA; Ala'a Elshaar<sup>1</sup>, MA; Brigid Dwyer<sup>2</sup>, MD; Ana Vives-Rodriguez<sup>1</sup>, MD; Rocco Palumbo<sup>1</sup>, PhD; Katherine Turk<sup>1,2,3</sup>, MD; Andrew Budson<sup>1,2,3</sup>, MD

<sup>1</sup>Center for Translational Cognitive Neuroscience, Veterans Affairs Boston Healthcare System, Boston, MA, United States

<sup>2</sup>Department of Neurology, Boston University School of Medicine, Boston, MA, United States

<sup>3</sup>Alzheimer's Disease Research Center, Boston University School of Medicine, Boston, MA, United States

**Corresponding Author:**

Anna Marin, BA

Center for Translational Cognitive Neuroscience

Veterans Affairs Boston Healthcare System

150 South Huntington Avenue

Boston, MA, 02130-4817

United States

Phone: 1 8573646184

Email: [amarin@bu.edu](mailto:amarin@bu.edu)

## Abstract

**Background:** Can home-based computerized cognitive training programs be a useful tool to sustain cognition and quality of life in patients with Alzheimer disease (AD)? To date, the progressive nature of the disease has made this question difficult to answer. Computerized platforms provide more accessibility to cognitive trainings; however, the feasibility of long-term, home-based computerized programs for patients with AD dementia remains unclear.

**Objective:** We aimed to investigate the feasibility of a 24-week home-based intervention program using the Constant Therapy app and its preliminary efficacy on cognition in patients with AD. Constant Therapy is a program developed for patients with speech and cognitive deficits. We hypothesized that patients with AD would use Constant Therapy daily over the course of the 24-week period.

**Methods:** Data were collected over a 48-week period. We recruited participants aged between 50 and 90 years with a diagnosis of mild cognitive impairment due to AD or mild AD dementia. Participants were randomly assigned to either the Constant Therapy (n=10) or active control (n=9) group. The Constant Therapy group completed a tablet-based training during the first 24 weeks; the second 24 weeks of computerized training were optional. The active control group completed paper-and-pencil games during the first 24 weeks and were invited to complete an optional Constant Therapy training during the second 24 weeks. Every 6 weeks, the participants completed the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). The participants independently accessed Constant Therapy using an Apple iPad. Our primary feasibility outcomes were the rate of adherence and daily use of Constant Therapy over 24 weeks. Our secondary outcomes were Constant Therapy performance over 24 weeks and change in RBANS scores between the 2 experimental groups.

**Results:** Feasibility analyses were computed for participants who completed 24 weeks of Constant Therapy. We found that long-term use of the Constant Therapy program was feasible in patients with AD over 24 weeks (adherence 80%; program use 121/168 days, for 32 minutes daily). These participants showed an overall improvement in accuracy and latency ( $P=.005$ ) in the Constant Therapy scores, as well as specific improvements in visual and auditory memory, attention, and arithmetic tasks. The Constant Therapy group showed improvement in the RBANS coding subtest. No unexpected problems or adverse events were observed.

**Conclusions:** Long-term (eg, 24 weeks) computerized cognitive training using Constant Therapy is feasible in patients with AD in the mild cognitive impairment and mild dementia stages. Patients adhered more to Constant Therapy than to the paper-and-pencil training over 24 weeks and improved their performance over time. These findings support the development of future randomized controlled trials that will investigate the efficacy of Constant Therapy to sustain cognitive function in patients with AD.

**Trial Registration:** ClinicalTrials.gov NCT02521558; <https://clinicaltrials.gov/ct2/show/NCT02521558>

**KEYWORDS**

cognitive training; Alzheimer disease dementia; technology

## Introduction

### Background

A total of 5.8 million people aged  $\geq 65$  years are living with Alzheimer disease (AD) dementia in the United States [1], highlighting a need for effective long-term cognitive interventions for these people. Cholinesterase inhibitors can help turn the clock back on the disease 6 to 12 months [2]. Aducanumab may possibly slow disease progression slightly, equivalent to 3 months, in patients with mild cognitive impairment (MCI) due to AD and mild AD dementia [3]. However, medications alone cannot halt the disease, and supplementing pharmacological interventions with nonpharmacological interventions has been shown to sustain cognition and quality of life more than medication alone [4-6].

Cognitive training programs are a traditional nonpharmacological intervention consisting of guided practice on standardized tasks to enhance specific cognitive functions, which may ultimately aid cognition and daily functioning [7]. During a typical session, patients complete tasks of varying difficulty, targeting different cognitive domains, such as memory, attention, and problem solving. The training is completed individually or in group sessions using paper and pencil or computerized programs under the supervision of a clinician. The repeated practice of tasks over time is aimed at improving or sustaining cognitive performance [7]. For example, patients with mild AD dementia showed improvements in their Mini-Mental State Examination scores when medication treatments were supplemented with a year of one-on-one regular cognitive training for 5 days weekly [5]. Supervised cognitive training has also been found to be beneficial in older adults with memory loss due to AD or vascular dementia by enhancing cognitive functioning and well-being in daily life [7-10]. Furthermore, a previous study of healthy older adults found that cognitive benefits were preserved 5 years after cognitive training [11]. Despite the potential benefits, traditional cognitive training programs require face-to-face contact, are expensive (staff prices from US \$15 to US \$100 per hour), and demand a significant time commitment (at least 60 minutes daily for 3 weeks) for the patient to make any gains. Therefore, it is challenging for patients with AD to adhere to traditional cognitive training programs [12].

Home-based, self-administered computerized cognitive training represents a practical alternative to overcome the expense and adherence challenges seen in traditional supervised-in-person cognitive training programs. Computerized cognitive training allows individuals to independently access cognitive exercises from their own computers, tablets, or other mobile devices at any time [13]. Home-based self-administered computerized cognitive training has been shown to benefit cognitive function as much as supervised-in-person training sessions in healthy older adults [12,14]. However, computerized cognitive training has produced mixed results in patients with AD. Some studies

show positive effects, others show a temporary effect or protection from decline, and some show no effect [15,16]. Two possible explanations for these discrepant findings could be the variability in the duration and level of difficulty of the training program. The progressive nature of AD-related cognitive decline also adds to the difficulty in accurately testing the effectiveness of home-based computerized programs. Many investigators endorse the need for research assessing the effects of longer and more individualized intervention programs that can adjust the level of task difficulty depending on the baseline cognitive function of patients [16-19].

Given the influence of factors such as age and clinical diagnosis on the effectiveness of computerized cognitive training on cognition [20], several platforms have been developed to provide more flexibility and accessibility for older adults [21], patients with AD [22], and other neurological diagnoses (eg, stroke, traumatic brain injury [TBI], and schizophrenia) [23].

We assessed the feasibility and preliminary efficacy of a 24-week individualized computerized program called Constant Therapy. Constant Therapy is a digitally delivered, cloud-based computerized training program developed for patients with speech and cognitive deficits. During a home-based, self-administered Constant Therapy session, patients practice computerized exercises in increasing order of difficulty. As they progress through their intervention schedule, the tasks change their level of difficulty, depending on the patients' individual progress. Constant Therapy allows patients to practice and advance independently such that patients experiencing different patterns of cognitive impairment can advance through the program at their own pace. Constant Therapy has been successfully implemented in individuals with aphasia due to stroke or TBI, with findings showing, on average, 70% compliance with the Constant Therapy self-administered training over the course of 10 to 20 weeks, improvements in task scores over time, and carryover to standardized assessment measures [24-26]. In these studies, accuracy (correct responses) and latency (reaction time) measures are used to quantify task performance. Increased accuracy and decreased latency characterize improved task performance [24]. The Constant Therapy platform has been studied both in the clinic and home environment. Patients using the platform at home make similar improvements compared with those who use the platform with their clinicians in the clinic [25].

No study has assessed the feasibility of the Constant Therapy app in the AD population. If feasible in the AD population, this type of home-based computerized intervention might have the potential to enhance cognitive functioning and support well-being in the daily lives of patients with AD.

Our study aimed to test the feasibility of long-term computerized cognitive Constant Therapy training (24 weeks) in the AD population.

## Objectives

The primary aim of this study was to assess the feasibility of Constant Therapy in patients with AD using a long-term individualized training program. We measured adherence to the Constant Therapy program over a period of 24 weeks using a randomized design. We hypothesized that patients with AD would use the Constant Therapy app daily and would adhere to the training over a 24-week period [27].

The secondary aim was to evaluate the preliminary efficacy of the training on the performance of Constant Therapy tasks and on standardized assessments of cognition and daily life functioning. As in the previous study by Des Roches et al [24], we assessed task performance (accuracy and latency), as any improvements could relate to willingness to continue adherence to the task and could suggest benefits to cognitive function. We hypothesized that the patients' performance after the long-term intervention period would be less impaired compared with patients who have not completed the Constant Therapy

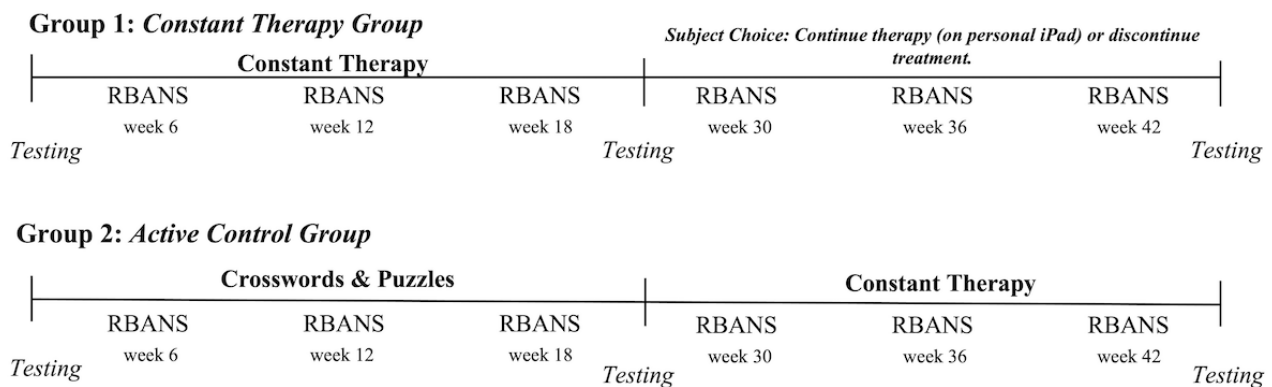
intervention. We also expected sustained improvement in performance on both the Constant Therapy tasks over 24 and 48 weeks and on neuropsychological scores after the intervention.

## Methods

### Study Design

The study was an unblinded randomized controlled trial (RCT) in which every newly recruited participant was randomly assigned to either training condition for 24 weeks. Participants' condition assignment was completed by the study coordinator using a web-based random number generator. Data were collected between October 2016 and January 2019. At the end of the 24 weeks, participants randomly assigned to the Constant Therapy training condition were allowed to continue using the app or to discontinue, while those in the active control training condition were offered the opportunity to use the app (48 weeks total; Figure 1).

**Figure 1.** Study design. RBANS: Repeatable Battery for the Assessment of Neuropsychological Status.



## Ethics Approval

All data from the patients' devices were anonymized during collection. This project was conducted under the Boston University Medical Center Institutional Review Board under study protocol H-34203.

## Participants

Patients were referred by the Boston University Alzheimer's Research Center or by practicing neuropsychologists, neurologists, or other associated discipline from Boston Medical Center. After referral, potential participants were contacted by a member of our research staff, who described the details of the study. We enrolled participants between 50 and 90 years of age that had a diagnosis of MCI due to AD or mild AD dementia. The diagnosis was confirmed by the referring clinician following the National Institutes of Aging-Alzheimer's Association criteria [28,29]. Moreover, the baseline cognitive battery was administered to all participants before beginning either training type to evaluate cognitive function at baseline and to ensure that the level of impairment did not exceed the mild AD dementia status.

Patients with any self-reported history of substance abuse, prior head trauma (eg, stroke or TBI), significant depression, or other mood disorders were not considered eligible to participate in

the study. Referring clinicians evaluated the exclusion of the patient from the study based on clinical notes and self-report.

## Study Training and Baseline Cognitive Testing

Before starting the intervention phase, participants were trained on how to use the Constant Therapy app and how to navigate an Apple iPad, such as how to access the app, switch between tasks, and complete each individual task. iPads were loaned to participants if they did not already own one. The cognitive testing and app training at the start of the study were done across 2 days for approximately 1 to 2 hours each day to avoid exhausting the participants. The research staff completed the 2-day preintervention training either from the laboratory site or in the participant's house, depending on the participant's preference. During the first day of training, the research staff completed baseline cognitive testing; on the second day, they completed an overview of the Constant Therapy platform and iPad (if assigned to the Constant Therapy condition) or an overview of the booklets containing the crosswords and puzzle (if assigned to the active control condition).

The baseline cognitive battery consisted of the Montreal Cognitive Assessment [30], the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), the Multifactorial Memory Questionnaire [31], an Activities of Daily Living Scale [32], a Quality of Life in Alzheimer's

Disease Scale [33], the Neuropsychological Assessment Battery for Memory [34], the Zarit Burden Interview [35], and the False Memory Questionnaire [36]. Following the initial neuropsychological assessment, the patients were randomly assigned to 1 of 2 groups (Figure 1). The same battery was repeated after 24 and 48 weeks of participation in the study to monitor changes in cognition and functioning in daily life throughout the study. To ensure consistency between testing sessions, the same research staff member completed assessments at month 0, month 24, and month 48. Testing sessions were conducted in the laboratory or at the participant's home. The RBANS was also administered at weeks 6, 12, 18, 30, 36, and 42 by the same assessor.

Research staff remotely monitored individual participant progress on the Constant Therapy platform daily. The study staff called the participants weekly to keep reminders consistent between both intervention groups. During these calls, the staff asked about adherence to the tasks, answered any questions, and reminded individuals to engage with the platform.

The research members completing the baseline, interim, and final assessments were not blinded to the participants' group assignment to ensure that participants could reach out for support and questions throughout the duration of the study.

## Intervention

### *Constant Therapy Group*

Patients in the Constant Therapy group (group 1, Figure 1) received the Constant Therapy program for a planned 24 weeks. Progress on the tasks was monitored daily, and weekly phone calls were completed to check in with participants. Participants were instructed to engage in Constant Therapy for approximately 30 minutes a day. The software recorded the amount of time spent performing the cognitive tasks. The neuropsychological testing battery performed at the start of the study was repeated at the end of the first 24 weeks. At this stage, participants were offered the option to either continue with the Constant Therapy training for an additional 24 weeks or to terminate their participation in the study. At the end of 48 weeks, the testing battery performed at the start of the study was repeated.

### *Active Control Group*

During the first 24 weeks of the study, participants in the active control group (group 2, Figure 1) received booklets containing

different types of puzzles and brain teasers (crossword puzzles, word search puzzles, Sudoku puzzles, and various types of math puzzles). They were instructed to perform these tasks for approximately 30 to 60 minutes per day. We monitored adherence to these puzzles weekly via phone conversations, mirroring the Constant Therapy group. After 24 weeks and the completion of the testing battery, the active control group was invited to participate in the Constant Therapy training for the following 24 weeks. The testing battery of neuropsychological assessments was completed at weeks 0, 24, and 48, as in the Constant Therapy group.

### **Constant Therapy Training Program**

Data were collected using the Constant Therapy app, which includes evidence-based speech, language, and cognitive exercises with varying levels of difficulty ranging from level 1 to level 10 (the software can be reviewed and accessed through the web link [37]). A total of 3 scores of 80% or higher advanced participants to the next difficulty level of a task. The tasks were designed with the aim of improving or stabilizing language, attention, and memory functioning. The exercises tested domains of language (naming, comprehension, speaking, reading, and writing) and cognitive skills (attention, executive skills and problem solving, mental flexibility, memory, and visuospatial skills). The Constant Therapy program recorded performance data (task accuracy and latency) as well as all other session activities (usability logs, use of built-in cues within the app, time stamps, and item completion indicators). For a more detailed review of the platform and cognitive tasks, refer to the study by Kiran et al [38].

In this pilot study, we examined the feasibility of Constant Therapy tasks in patients with AD, with the future goal of testing in a larger RCT how Constant Therapy tasks could help stabilize or improve some of the cognitive domains most prominently affected by AD (eg, memory, language, and executive function) [39,40]. All tasks were self-paced and self-administered by the participants. All tasks gave the participants the option to skip or quit at any time if they felt fatigued or frustrated. The Constant Therapy tasks used in this study were not modified for the population with AD.

### **Outcome Measures**

We collected primary feasibility measures and secondary preliminary efficacy measures (Textbox 1).

**Textbox 1.** Outcome measures.**Outcome measures**

## 1. Primary feasibility measures

- Overall adherence to the study design up to 24 weeks
- Adherence rates in the Constant Therapy and the active control training during the first 24 weeks
- Use of the Constant Therapy app over the first 24 weeks
- Any engagement with the app during voluntary continuation to 48 weeks

## 2. Secondary preliminary efficacy variables

- Constant Therapy tasks performance (accuracy and latency):
  - Arithmetic tasks (addition, multiplication, subtraction, and division)
  - Auditory tasks (environmental sound matching, spoken word matching, voicemail, and auditory command)
  - Visual tasks (calendar reading, clock math, clock reading, map reading, mental rotation, pattern recreation, picture matching, face matching, picture n-back memory, playing-card slapjack, symbol matching, written word matching, and flanker)
  - Quantitative reasoning tasks (currency, functional math, number pattern, and word problem)
- Repeatable Battery for the Assessment of Neuropsychological Status subtest scores at week 0, 6, 12, 18, and 24:
  - Memory: list learning, list recall, list recognition, story immediate recall, story, and delayed recall
  - Language: picture naming and semantic fluency
  - Executive function: digit span and coding
  - Visuospatial and constructional: figure copy and line orientation

**Statistical Analysis****Sample Size**

Prior work has studied the feasibility and preliminary efficacy of computerized cognitive training over 12 weeks in patients with MCI and AD dementia using RCTs with sample sizes of 11 [41], 20 [42], and 22 participants [43].

**Analytic Plan**

The overall data analyses conducted aimed to primarily assess the feasibility of the Constant Therapy program in patients with AD over the course of 24 and 48 weeks, with secondary analyses examining the preliminary efficacy of the Constant Therapy program for improving or stabilizing cognitive function over an extended period.

**Demographics**

A total of 19 participants (18, 85% male and 1, 5% female) aged 64 to 85 years from the Boston University Alzheimer's Disease Research Center and Boston Medical Center were enrolled in the study and were randomly assigned to an experimental condition. They met the criteria for MCI due to AD (n=7, 37%) or mild AD dementia (n=12, 63%), as described by the National Institutes of Aging-Alzheimer's Association criteria [28,29]. All study participants were non-Latino White people.

The demographics of the 19 participants were analyzed using descriptive statistics and included age, education (years of schooling), and baseline scores of cognition and daily life functioning. We used the Mann-Whitney Wilcoxon test to

evaluate differences between the Constant Therapy and active control group.

**Study Adherence**

We measured the rate of adherence to this novel long-term intervention program lasting between 24 and 48 weeks. Adherence rate was calculated by counting the number of participants enrolled in the program every 6 weeks. As individuals were given the option to continue after 24 weeks, we also reported the rate of adherence up to 48 weeks.

**Constant Therapy Usage**

We then performed an analysis of engagement to the Constant Therapy program by computing the average number of days each participant spent on the app over the course of the intervention, as well as the average time spent on the app daily.

**Constant Therapy Tasks Performance**

To assess our secondary aim, we analyzed the change in performance on the Constant Therapy tasks measured by accuracy and latency scores for each task. We conducted a Wilcoxon signed-rank test to determine changes in performance (accuracy and latency) for each task. We compared scores for each task at the start (average of the first 10 trials) and at the end (average of the last 10 trials) of the 24 weeks of training. The first 3 observations were excluded before averaging the 10 initial scores to account for the practice time required to adapt to the Constant Therapy platform [24].

### ***Preliminary Efficacy of Constant Therapy on Clinical Variables***

Our ability to compare changes in cognitive performance between week 0 and week 24 was limited. The RBANS was the only measure that was repeated 5 times throughout the 24-week period (baseline, week 6, week 12, week 18, and week 24). The rest of the outcome measures (Montreal Cognitive Assessment, Zarit Burden Interview, Multifactorial Memory Questionnaire, Activities of Daily Living Scale, Quality of Life in Alzheimer's Disease Scale, False Memory Questionnaire, and Neuropsychological Assessment Battery) were administered twice (baseline and week 24) and could not be further analyzed due to a higher dropout rate in the active control group than in the Constant Therapy group. Thus, we conducted an exploratory analysis using only the RBANS subtest scores to evaluate the preliminary efficacy of the Constant Therapy training. None of the participants dropped out from week 0 to week 6; therefore, we assessed the changes in performance over the first 6 weeks

between the Constant Therapy and active control groups. While normality was not always violated (only in 4 out of 12 RBANS submeasures), we computed a Mann-Whitney test to protect against type 1 and type 2 errors that are likely to occur with our small sample size. To further explore the feasibility of the training, for the RBANS subtest that were normally distributed, we performed a 2×5 repeated measures ANOVA with the factors of group (2: Constant Therapy vs active control) and time (5: week 0, 6, 12, 18, and 24). This was tested for group differences at each time point. Post hoc comparisons were performed using the Tukey honestly significant difference test.

## ***Results***

---

### **Demographics**

The Constant Therapy and active control groups did not differ in demographic variables, as determined using the Mann-Whitney Wilcoxon test (Table 1).

**Table 1.** Demographics at baseline test.

Characteristics	Constant Therapy (n=10)	Active control (n=9)	P value
Sex (male), n (%)	10 (100)	8 (89)	N/A <sup>a</sup>
<b>Diagnosis, n (%)</b>			
MCI <sup>b</sup> due to AD <sup>c</sup>	4 (40)	3 (33)	N/A
Mild AD dementia	6 (60)	6 (67)	N/A
Ethnicity and race (non-Latino White)	10 (100)	9 (100)	N/A
Age (years), median (IQR)	72.50 (14.00) <sup>d</sup>	75.00 (13.00)	.78
Education, median (IQR)	14.00 (4.00)	14.00 (5.50)	.91
<b>Outcome measures, median (IQR)</b>			
MoCA <sup>e</sup>	20.50 (6.00) <sup>d</sup>	21.50 (2.00) <sup>d</sup>	.97
ZBI <sup>f</sup>	13.00 (15.25)	6.00 (10.00) <sup>d</sup>	.97
MMQ <sup>g</sup>	103.00 (26.75) <sup>d</sup>	129.00 (29.00) <sup>d</sup>	.10
ADL <sup>h</sup>	12.50 (22.75)	26.00 (39.00) <sup>d</sup>	.50
QOL-AD <sup>i</sup>	37.50 (19.75) <sup>d</sup>	36.00 (14.00) <sup>d</sup>	.91
FMQ <sup>j</sup>	43.50 (45.75) <sup>d</sup>	56.00 (47.00) <sup>d</sup>	.84
<b>RBANS<sup>k</sup>, median (IQR)</b>			
Total list learning	20.00 (9.00) <sup>d</sup>	20.50 (8.50) <sup>d</sup>	.66
List recall	1.00 (3.25)	0.50 (2.75)	.45
List recognition	15.00 (5.25) <sup>d</sup>	15.00 (4.75) <sup>d</sup>	.91
Story memory (immediate recall)	14.00 (4.50)	10.50 (4.00) <sup>d</sup>	.11
Story memory (delayed recall)	3.50 (5.25) <sup>d</sup>	4.00 (4.00) <sup>d</sup>	.72
Digit span	9.500 (5.50) <sup>d</sup>	10.00 (3.00) <sup>d</sup>	.72
Figure copy	20.00 (2.00)	19.00 (1.00)	.11
Figure recall	3.500 (13.50)	4.00 (6.00)	.66
Semantic fluency	15.00 (10.25)	16.50 (6.75) <sup>d</sup>	.36
Line orientation	17.00 (5.25) <sup>d</sup>	17.00 (5.25) <sup>d</sup>	.50
Picture naming	9.50 (1.00)	9.50 (1.00)	.99
Coding	27.00 (14.75) <sup>d</sup>	32.50 (12.50) <sup>d</sup>	.72

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

<sup>b</sup>MCI: mild cognitive impairment.

<sup>c</sup>AD: Alzheimer disease.

<sup>d</sup>Normally distributed.

<sup>e</sup>MoCA: Montreal Cognitive Assessment.

<sup>f</sup>ZBI: Zarit Burden Interview.

<sup>g</sup>MMQ: Multifactorial Memory Questionnaire.

<sup>h</sup>ADL: Activities of Daily Living Scale.

<sup>i</sup>QOL-AD: Quality of Life in Alzheimer's Disease Scale.

<sup>j</sup>FMQ: False Memory Questionnaire.

<sup>k</sup>RBANS: Repeatable Battery for the Assessment of Neuropsychological Status.



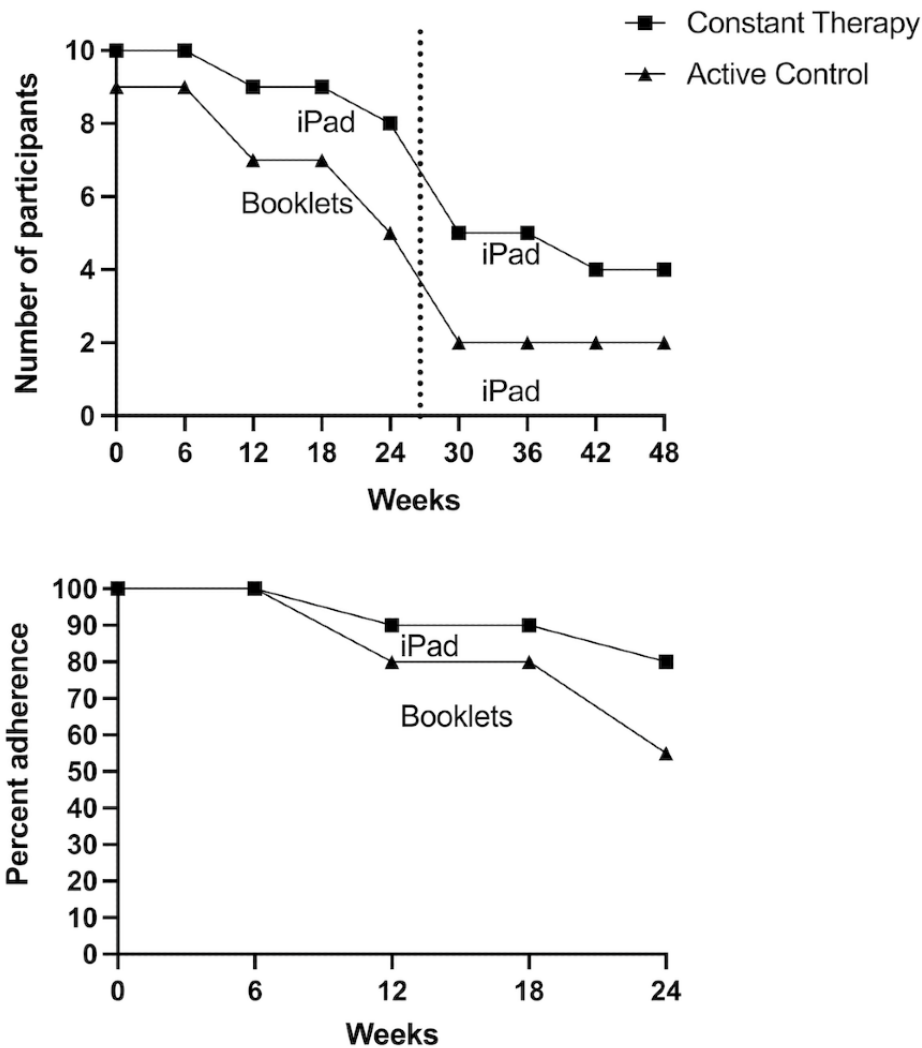
**Study Adherence**

**Constant Therapy Group Adherence**

As shown in Figure 2, 80% (8/10) of participants in the Constant

Therapy group completed 24 weeks of the intervention; 5 patients continued the study beyond the 24-week period, 1 to 42 weeks (10%) and 4 to 48 weeks (40%).

Figure 2. Study adherence.



**Active Control Group Adherence**

In total, 55% (5/9) of participants in the active control group completed the 24 weeks of the study. Overall, 22% (2/9) of participants elected to engage with the Constant Therapy app until week 48. No unexpected problems were observed in this study. The reasons for dropping out of the intervention were not collected, and individuals who dropped out of the intervention were also discontinued from their postintervention clinical assessment.

The following analysis used data collected between week 0 and week 24.

**Constant Therapy Use**

We examined app use over 24 weeks (168 days) in the Constant Therapy intervention group (Multimedia Appendix 1). On average, participants engaged with the app a mean of 121.4 (SD 38.56, 95% CI 97.50-145.30) days, with 31.70 (SD 9.94, 95%

CI 25.54-37.86) minutes spent on the app per day. Participants with MCI due to AD (n=4) spent a mean of 147.50 (SD 21.30, 95% CI 126.63-168.37) days, with 38.16 (SD 11.09, 95% CI 27.29-49.02) minutes per day on the app and participants with mild AD dementia (n=6) spent a mean of 104 (SD 38.67, 95% CI 73.05-134.94) days, with 27.39 (SD 6.97, 95% CI 21.82-32.97) minutes per day on the app.

**Performance on Constant Therapy Tasks**

To assess the preliminary efficacy of the intervention, we analyzed the accuracy and latency scores of the Constant Therapy group at the start and end of the 24-week Constant Therapy training (Table 2). Table 2 shows the task scores that improved over 24 weeks. In the Multimedia Appendices 2 and 3 we also present the by-participant task progression over the 24-week intervention period (Multimedia Appendix 1) and by participant improvements and decrements across the individual tasks (Multimedia Appendix 2).

**Table 2.** Constant Therapy tasks performance over 24 weeks.

Task	Start, median (IQR)	End, median (IQR)	z-score	P value
<b>Accuracy</b>				
Cumulative	0.89 (0.04)	0.91 (0.02)	2.8	.005
Addition	0.93 (0.08)	0.98 (0.05)	2.49	.01
Environmental sound matching	0.86 (0.14)	0.95 (0.07)	2.55	.01
Picture N-back memory	0.88 (0.18)	0.95 (0.09)	2.93	.02
Written word matching	0.87 (0.07)	0.90 (0.06)	1.99	.047
<b>Latency</b>				
Cumulative	25.78 (11.57)	21.96 (9.6)	-2.8	.005
Addition	17.88 (6.67)	12.37 (4.34)	-2.8	.005
Environmental sound matching	33.52 (16.13)	25.45 (10.4)	-2.67	.008
Picture N-back memory	25.75 (3.44)	24.3 (3.12)	-2.8	.005
Auditory Command	34.8 (27.51)	30.02 (10.22)	-2.8	.005
Calendar reading	19.27 (15.99)	15.74 (14.31)	-2.2	.03
Clock reading	6.22 (3.27)	4.82 (5.09)	-2.07	.04
Currency	25.85 (19.23)	21.01 (11.81)	-2.39	.02
Picture matching	50.28 (28.03)	47.41 (30.91)	-2.39	.02
Spoken Word matching	72.58 (68.21)	68.86 (56.2)	-1.99	.047
Playing-card slapjack	22.8 (1.75)	21.26 (0.99)	-2.07	.04
Symbol matching	18.79 (10.87)	15.56 (10.31)	-2.08	.005
Voicemail	29.99 (6.64)	28.3 (4.9)	-1.99	.046
Word problem	62.52 (59.5)	49.68 (41.25)	-2.5	.01

### ***Preliminary Efficacy of Constant Therapy on Clinical Variables***

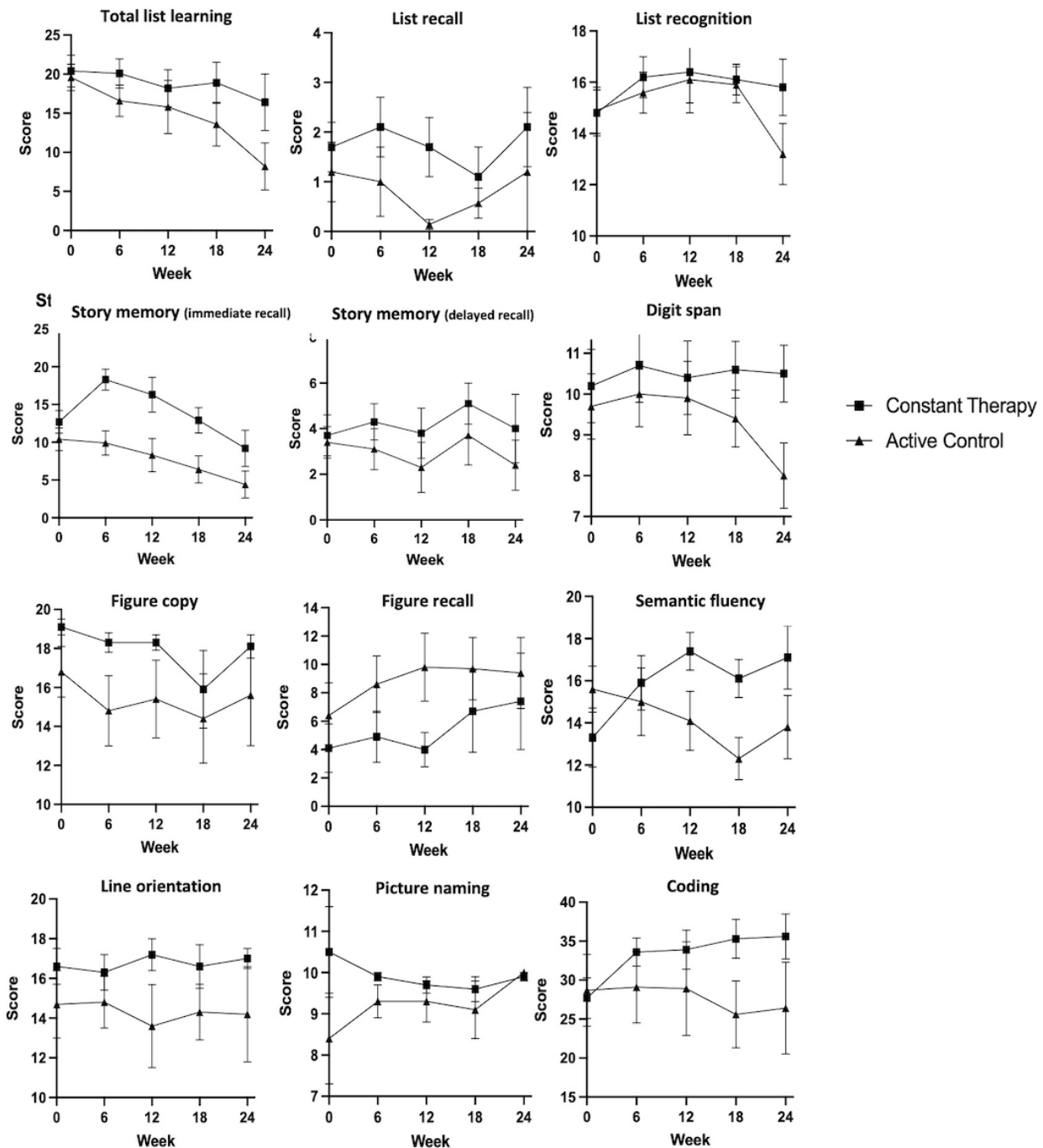
As stated in our analytic plan, we compared changes from week 0 to week 6 using the Mann-Whitney test for all RBANS subtests. When computing Mann-Whitney, changes in RBANS coding scores from week 0 to week 6 were significantly different between the 2 condition groups ( $U=80.00$ ;  $z=2.867$ ;  $P=.003$ ), and the Constant Therapy group, with a median of 6.5 (IQR 5.75), had a larger improvement in coding scores compared with the active control group, with a median of 1 (IQR 3) after the first 6 weeks in the study. No other differences were observed between the RBANS subscores between weeks 0 and 6 ( $P>.139$ ).

Next, since normally distributed, we also conducted a repeated measures ANOVA to explore changes in the coding scores over the full 24 weeks. We found a main effect of time ( $F_{4,32}=4.34$ ;

$P=.006$ ;  $\eta^2=0.35$ ). Coding score at week 0 (mean 24.71, SD 9.08; SE 3.18) was higher than (1) week 6 (mean 29.02, SD 10.60; SE 3.32;  $P=.002$ ); (2) week 12 (mean 29.36, SD 11.27; SE 3.46;  $P=.01$ ); (3) week 18 (mean 28.69, SD 11.73; SE 3.50;  $P=.02$ ); and (4) week 24 (mean 27.05, SD 11.47; SE 3.29;  $P=.02$ ). No main effect of group was found ( $F_{1,8}=3.22$ ;  $P=.11$ ;  $\eta^2=0.29$ ). There was an interaction between group and time ( $F_{4,32}=4.06$ ;  $P=.009$ ;  $\eta^2=0.34$ ). The Constant Therapy group performed better at week 6 (mean 33.60, SD 6.26; SE 3.63;  $P=.001$ ), week 12 (mean 33.89, SD 7.59; SE 3.79;  $P=.006$ ), week 18 (mean 35.33, SD 7.70; SE 3.84;  $P=.005$ ) and week 24 (mean 35.62, SD 8.18; SE 3.60;  $P<.001$ ) compared with week 0 (mean 27.7, SD 7.00; SE 3.48). No significant differences were observed in the active control group over time.

The score changes across the first 24 weeks in the Constant Therapy and active control groups in each RBANS subtest administered are presented in [Figure 3](#).

**Figure 3.** Mean and SE for the 12 Repeatable Battery for the Assessment of Neuropsychological Status subtests over 24 weeks.



## Discussion

### Principal Findings

Our study aimed to examine the feasibility of home-based, self-administered, and long-term individualized cognitive training using Constant Therapy in patients with AD. In addition, as a secondary aim, we sought to evaluate the preliminary efficacy of the Constant Therapy training program on neuropsychological performance. Overall, this feasibility study aims to inform the development of a future RCT.

We predicted that patients with AD would adhere to the training over a 24-week period using the Constant Therapy app.

Consistent with this prediction, we found that long-term use of the Constant Therapy program was feasible in a patient population with AD, as shown by the rate of adherence (80%) and use of the program (average of 121 out of 168 days for 32 minutes daily) over 24 weeks. In comparison, the participants assigned to the active control group had a 55% adherence rate to the study at 24 weeks. Our adherence rates are comparable with those of other RCTs testing computerized cognitive training in older adults with cognitive impairment (eg, 77% of patients with MCI in a 6-week period study adhered to the intervention and 76% adhered to the control sessions [44]), in patients with other neurological diseases (eg, 76% adherence to an 18-week training period in patients with Huntington disease [45]), and

in more heterogeneous cohorts (eg, 83% adherence over a 24-week training period in patients with a variety of neurological and psychiatric diseases [46]).

Adherence rates dropped during the second optional part of the study, both in the Constant Therapy group (40%), who voluntarily continued the training, and the active control group (22%), who could begin the Constant Therapy training at 24 weeks. Due to the drop in adherence and loss of power between the first 24 weeks and the second 24 weeks, we did not examine the data collected between weeks 24 and 48. Although the data were not assessed, low adherence to the second portion of the study is helpful in informing the design and length of future RCTs. A future RCT should focus on testing the efficacy of a 24-week intervention training program.

The overall sustained performance on computerized tasks shows that the individualized training approach modeled by Constant Therapy is appropriate for a population with AD. Furthermore, not only did patients sustain their performance in the tasks but they also displayed improvements in performance that are important to report for future RCT studies. Specifically, patients performed more accurately over time in the task training domains of visual and auditory memory, attention, and arithmetic. Latency also improved in tasks related to visuospatial processing, visual and auditory memory, attention, quantitative reasoning, and arithmetic skills. Faster reaction time in the Constant Therapy tasks may suggest improvement in processing speed, as well as improved adaptability to computerized tasks. Faster reaction time not paired with improved accuracy may also represent increased disinhibition while completing the task. These positive findings suggest that the 24-week intervention program is feasible and underscore the need for future larger studies to test the effectiveness of Constant Therapy as a long-term training program for patients with AD dementia.

We also predicted that gains made using the Constant Therapy program would transfer to neuropsychological test performance. We monitored changes in performance by administering the RBANS every 6 weeks and the Constant Therapy group showed an improvement in coding abilities over time. Coding performance improved during the first 6 weeks and then remained stable over the remaining 18 weeks. Although we experienced a higher drop in adherence in the active control group compared with the Constant Therapy group over the first 24 weeks (which ultimately limits our ability to interpret findings pertaining to this prediction), the observed change in performance between the Constant Therapy and active control groups on the RBANS subtest of coding, despite the small numbers of participants, indicate that, to some extent, gains made during Constant Therapy have the potential to transfer onto neuropsychological test performance.

Coding was the only measure that showed an improvement in performance over time when using Constant Therapy. Coding is often used as a measure of executive function in the neuropsychological assessment of patients with dementia [47,48]. This pattern of results is consistent with previous literature showing that computerized cognitive training may lead to improvements in cognitive performance in the executive function domain [49]. While exploratory in nature, these results

support previous work showing that computerized cognitive training programs have the potential to improve performance in neuropsychological tests and help mitigate cognitive decline in older adults with AD [15]. The results of the outcome measures in this study highlight the usefulness of frequent neuropsychological monitoring when designing long-term intervention studies. Future research examining the feasibility of long-term individualized computerized programs in similar populations could incorporate frequent neuropsychological tests into their designs to better assess the impact of these interventions on cognition.

Long-term computerized cognitive trainings (eg, 24 weeks) have shown to be effective in healthy older adults [16]. Our primary aim was to examine whether they are feasible in the population with AD using the Constant Therapy platform. Although several studies have investigated computerized cognitive training in patients with AD, most have consisted of intervention periods that do not exceed 8 weeks [44]. In addition, until now, Constant Therapy has most often been used for the rehabilitation of language or cognitive deficits caused by stroke or TBI [24,50] and has not yet been tested in patients in AD. This study indicates that the Constant Therapy home-based individualized program is feasible for 24 weeks and may be a beneficial tool for patients with AD. These findings build on recent evidence showing that patients with dementia demonstrate improvement in global cognitive function when provided with individualized cognitive training in both a traditional or remote clinical setting [51].

Furthermore, our data indicate that the Constant Therapy program can be feasible in both the MCI and mild dementia stages of AD, providing supporting evidence that the individualized progression in task difficulty and the length of the intervention were acceptable to patients given their adherence to the platform. A future larger RCT is needed to examine how computerized training may impact cognition and function in the daily lives of patients with AD. Finally, we note that the COVID-19 pandemic has posed new challenges to the feasibility of in-person cognitive training. Thus, now more than ever, the investigation of home-based, self-administered computerized platforms is essential to assist older adults with AD and related cognitive disorders [52].

## Limitations

This study tests the feasibility and attempts to test the preliminary efficacy of individualized, long-term, home-based computerized cognitive therapy in patients with AD. A limitation of our study is that we could not examine the data collected from weeks 24 to 48 due to the low rate of participation beyond 24 weeks. Nevertheless, the training length we examined (24 weeks) is still a valuable strength of our study, as it is a longer training period than that commonly tested in patients with AD. Finding new ways to actively involve caregivers in the program as a source of support throughout the intervention may help increase overall adherence over long periods. In addition, the drop in adherence over time in the active control group was greater than that in the Constant Therapy group. We believe that the use of paper-and-pencil games might have influenced retention in the active control group. While previous studies

have shown no significant difference between the use of paper and computerized active control tasks on training effect [53], most of the work showing positive benefits from cognitive training used a computerized active control condition [16]. Computerized tasks are more interactive and entertaining, which could result in increased levels of motivation for the active control group [12].

Some limitations due to practical implementation considerations, the study design, and the patient population must be acknowledged. First, study staff members administering the batteries were not blinded, and the total sample size was 19. While this choice enabled our staff to answer questions related to the app during weekly check-ins if needed, this study feature, in combination with the small number of participants that engaged with the app, limited the conclusions. As such, we consider the preliminary efficacy findings to be exploratory. Nevertheless, our findings, especially regarding feasibility, help inform the suitability of training for larger trials in the future. In addition, we administered only the limited testing battery (ie, RBANS) at each 6-week interval and not surveys measuring metamemory, quality of life, and caregiver burden, for example. While this choice allowed us to reduce participant burden (because participants only had to complete approximately 40 minutes of testing), it is unknown whether engagement with the app benefited some other dimension of the participants' and caregivers' life experiences other than patient cognition. Furthermore, acceptability measures, such as satisfaction levels regarding the study in general, its individual components, or open-ended questions to collect feedback, were not assessed. Future RCTs should assess acceptability measures to help improve users' experiences and overall adherence to interventions. In addition, larger RCTs should monitor and report the time point of the study where each participant dropped out. This categorization is important to examine the efficacy of the training in relation to the frequency and length of the assessments over the study period.

Finally, we acknowledge that we did not formally measure why participants stopped engaging with the app. However, in our discussions with the study coordinator, some participants may have experienced some difficulty with the auditory tasks, especially when the level of difficulty increased. It is possible that hearing difficulty might have impacted adherence to the training program as a whole and interfered with the effectiveness of the intervention on cognitive function.

Finally, due to the homogeneity of our sample, we cannot make any conclusions about the feasibility of the Constant Therapy training in other demographic groups. Of note, a factor contributing to this limitation is that our study lacked representation of women, and racially and ethnically minoritized groups.

The implications of this limitation are well acknowledged in such cognitive rehabilitation studies. For example, motivation, initial cognitive ability, and income level have the potential to affect cognitive interventions [54]. Furthermore, the use of technology may be differentially accessed and used based on factors such as one's level of education [55]. While published work has shown that Constant Therapy rehabilitation training is equally feasible for individuals with speech language disorders who live in different geographic areas (urban vs rural) [26], future RCT studies should further explore the role of different demographic factors, including ethnicity, socioeconomic status, education, and gender, on the feasibility of Constant Therapy in the AD population.

## Conclusions

Home-based, self-administered, computerized cognitive trainings are a potential tool to help sustain cognitive function in patients with dementia [15,51]. Our study aimed to test the feasibility and preliminary efficacy of the Constant Therapy platform for patients with AD. Despite some of the challenging aspects of long-term intervention studies, our data show that Constant Therapy is a feasible platform for patients with AD for an intervention period of 24 weeks. Our findings support previous evidence that home-based self-administered Constant Therapy programs are a feasible alternative to in-person supervised cognitive training programs [25]. The patients with AD in our study engaged in Constant Therapy tasks and improved their performance over time. An exploratory analysis also showed promising changes in the RBANS coding subtest, which is a measure of executive function.

Future trials are necessary to investigate the efficacy of Constant Therapy training over a 24-week period on cognition and daily life function and to test adherence to longer intervention periods, such as 48 weeks. Thus, while long-term individualized Constant Therapy is feasible in patients with AD, more research is needed to explore its benefits and the factors that can influence its efficacy.

---

## Acknowledgments

This study was funded by the Wallace Coulter Foundation (grant 55200086) and Boston University Alzheimer's Disease Research Center (grant P30 AG072978).

---

## Authors' Contributions

AM was responsible for data analysis and interpretation as well as manuscript writing. KS contributed to the interpretation of the data analysis and manuscript writing. AE and BD contributed to study design, conceptualization, and data collection. AV-R contributed to the data analysis and interpretation. RP and RD contributed to data analysis and statistical support. KT contributed to the study design and conceptualization as well as the interpretation of data analysis. AB contributed to the study design and conceptualization, interpretation of data analysis, and edited the manuscript.

---

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Constant Therapy use over the first 24 weeks.

[[DOCX File , 153 KB - formative\\_v6i9e34450\\_app1.docx](#) ]

### Multimedia Appendix 2

Constant Therapy task levels completion over 24 weeks.

[[DOCX File , 24 KB - formative\\_v6i9e34450\\_app2.docx](#) ]

### Multimedia Appendix 3

Constant Therapy tasks performance over 24 weeks.

[[DOCX File , 50 KB - formative\\_v6i9e34450\\_app3.docx](#) ]

### Multimedia Appendix 4

CONSORT-EHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 2755 KB - formative\\_v6i9e34450\\_app4.pdf](#) ]

## References

1. Rajan K, Weuve J, Barnes L, McAninch E, Wilson R, Evans D. Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020-2060). *Alzheimers Dement* 2021 Dec;17(12):1966-1975 [[FREE Full text](#)] [doi: [10.1002/alz.12362](https://doi.org/10.1002/alz.12362)] [Medline: [34043283](https://pubmed.ncbi.nlm.nih.gov/34043283/)]
2. Budson AE, Solomon PR. *Memory Loss, Alzheimer's Disease, and Dementia A Practical Guide for Clinicians*. 3rd edition. Amsterdam, The Netherlands: Elsevier Health Sciences; 2021.
3. Cummings J, Aisen P, Lemere C, Atri A, Sabbagh M, Salloway S. Aducanumab produced a clinically meaningful benefit in association with amyloid lowering. *Alzheimers Res Ther* 2021 May 10;13(1):98 [[FREE Full text](#)] [doi: [10.1186/s13195-021-00838-z](https://doi.org/10.1186/s13195-021-00838-z)] [Medline: [33971962](https://pubmed.ncbi.nlm.nih.gov/33971962/)]
4. Requena C, Maestú F, Campo P, Fernández A, Ortiz T. Effects of cholinergic drugs and cognitive training on dementia: 2-year follow-up. *Dement Geriatr Cogn Disord* 2006;22(4):339-345. [doi: [10.1159/000095600](https://doi.org/10.1159/000095600)] [Medline: [16954689](https://pubmed.ncbi.nlm.nih.gov/16954689/)]
5. Requena C, López Ibor MI, Maestú F, Campo P, López Ibor JJ, Ortiz T. Effects of cholinergic drugs and cognitive training on dementia. *Dement Geriatr Cogn Disord* 2004;18(1):50-54. [doi: [10.1159/000077735](https://doi.org/10.1159/000077735)] [Medline: [15084794](https://pubmed.ncbi.nlm.nih.gov/15084794/)]
6. Rozzini L, Costardi D, Chilovi BV, Franzoni S, Trabucchi M, Padovani A. Efficacy of cognitive rehabilitation in patients with mild cognitive impairment treated with cholinesterase inhibitors. *Int J Geriatr Psychiatry* 2007 Apr;22(4):356-360. [doi: [10.1002/gps.1681](https://doi.org/10.1002/gps.1681)] [Medline: [17117398](https://pubmed.ncbi.nlm.nih.gov/17117398/)]
7. Bahar-Fuchs A, Clare L, Woods B. Cognitive training and cognitive rehabilitation for mild to moderate Alzheimer's disease and vascular dementia. *Cochrane Database Syst Rev* 2013 Jun 05;2013(6):CD003260 [[FREE Full text](#)] [doi: [10.1002/14651858.CD003260.pub2](https://doi.org/10.1002/14651858.CD003260.pub2)] [Medline: [23740535](https://pubmed.ncbi.nlm.nih.gov/23740535/)]
8. Fernández-Prado S, Conlon S, Mayán-Santos JM, Gandoy-Crego M. The influence of a cognitive stimulation program on the quality of life perception among the elderly. *Arch Gerontol Geriatr* 2012;54(1):181-184. [doi: [10.1016/j.archger.2011.03.003](https://doi.org/10.1016/j.archger.2011.03.003)] [Medline: [21458869](https://pubmed.ncbi.nlm.nih.gov/21458869/)]
9. Trebbastoni A, Imbriano L, Podda L, Rendace L, Sacchetti ML, Campanelli A, et al. Cognitive training in patients with Alzheimer's disease: findings of a 12-month randomized controlled trial. *Curr Alzheimer Res* 2018 Mar 14;15(5):452-461. [doi: [10.2174/1567205014666171113105044](https://doi.org/10.2174/1567205014666171113105044)] [Medline: [29141548](https://pubmed.ncbi.nlm.nih.gov/29141548/)]
10. Nousia A, Siokas V, Aretouli E, Messinis L, Aloizou A, Martzoukou M, et al. Beneficial effect of multidomain cognitive training on the neuropsychological performance of patients with early-stage Alzheimer's disease. *Neural Plast* 2018;2018:2845176 [[FREE Full text](#)] [doi: [10.1155/2018/2845176](https://doi.org/10.1155/2018/2845176)] [Medline: [30123243](https://pubmed.ncbi.nlm.nih.gov/30123243/)]
11. Willis SL, Tennstedt SL, Marsiske M, Ball K, Elias J, Koepke KM, ACTIVE Study Group. Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA* 2006 Dec 20;296(23):2805-2814 [[FREE Full text](#)] [doi: [10.1001/jama.296.23.2805](https://doi.org/10.1001/jama.296.23.2805)] [Medline: [17179457](https://pubmed.ncbi.nlm.nih.gov/17179457/)]
12. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. *PLoS One* 2012;7(7):e40588 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0040588](https://doi.org/10.1371/journal.pone.0040588)] [Medline: [22792378](https://pubmed.ncbi.nlm.nih.gov/22792378/)]
13. Harvey PD, McGurk SR, Mahncke H, Wykes T. Controversies in computerized cognitive training. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2018 Nov;3(11):907-915 [[FREE Full text](#)] [doi: [10.1016/j.bpsc.2018.06.008](https://doi.org/10.1016/j.bpsc.2018.06.008)] [Medline: [30197048](https://pubmed.ncbi.nlm.nih.gov/30197048/)]
14. Zokaei N, MacKellar C, Čepukaitytė G, Patai EZ, Nobre AC. Cognitive training in the elderly: bottlenecks and new avenues. *J Cogn Neurosci* 2017 Sep;29(9):1473-1482. [doi: [10.1162/jocn\\_a\\_01080](https://doi.org/10.1162/jocn_a_01080)] [Medline: [27897677](https://pubmed.ncbi.nlm.nih.gov/27897677/)]

15. Klimova B, Maresova P. Computer-based training programs for older people with mild cognitive impairment and/or dementia. *Front Hum Neurosci* 2017 May 16;11:262 [FREE Full text] [doi: [10.3389/fnhum.2017.00262](https://doi.org/10.3389/fnhum.2017.00262)] [Medline: [28559806](https://pubmed.ncbi.nlm.nih.gov/28559806/)]
16. Bozoki A, Radovanovic M, Winn B, Heeter C, Anthony JC. Effects of a computer-based cognitive exercise program on age-related cognitive decline. *Arch Gerontol Geriatr* 2013;57(1):1-7. [doi: [10.1016/j.archger.2013.02.009](https://doi.org/10.1016/j.archger.2013.02.009)] [Medline: [23542053](https://pubmed.ncbi.nlm.nih.gov/23542053/)]
17. Shao Y, Mang J, Li P, Wang J, Deng T, Xu Z. Computer-based cognitive programs for improvement of memory, processing speed and executive function during age-related cognitive decline: a meta-analysis. *PLoS One* 2015 Jun 22;10(6):e0130831 [FREE Full text] [doi: [10.1371/journal.pone.0130831](https://doi.org/10.1371/journal.pone.0130831)] [Medline: [26098943](https://pubmed.ncbi.nlm.nih.gov/26098943/)]
18. Ten Brinke LF, Best JR, Crockett RA, Liu-Ambrose T. The effects of an 8-week computerized cognitive training program in older adults: a study protocol for a randomized controlled trial. *BMC Geriatr* 2018 Jan 30;18(1):31 [FREE Full text] [doi: [10.1186/s12877-018-0730-6](https://doi.org/10.1186/s12877-018-0730-6)] [Medline: [29378515](https://pubmed.ncbi.nlm.nih.gov/29378515/)]
19. Belleville S. Cognitive training for persons with mild cognitive impairment. *Int Psychogeriatr* 2008 Feb;20(1):57-66. [doi: [10.1017/S104161020700631X](https://doi.org/10.1017/S104161020700631X)] [Medline: [17958927](https://pubmed.ncbi.nlm.nih.gov/17958927/)]
20. Sternberg DA, Ballard K, Hardy JL, Katz B, Doraiswamy PM, Scanlon M. The largest human cognitive performance dataset reveals insights into the effects of lifestyle factors and aging. *Front Hum Neurosci* 2013;7:292 [FREE Full text] [doi: [10.3389/fnhum.2013.00292](https://doi.org/10.3389/fnhum.2013.00292)] [Medline: [23801955](https://pubmed.ncbi.nlm.nih.gov/23801955/)]
21. Rebok GW, Ball K, Guey LT, Jones RN, Kim H, King JW, ACTIVE Study Group. Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. *J Am Geriatr Soc* 2014 Jan;62(1):16-24 [FREE Full text] [doi: [10.1111/jgs.12607](https://doi.org/10.1111/jgs.12607)] [Medline: [24417410](https://pubmed.ncbi.nlm.nih.gov/24417410/)]
22. Fiatarone Singh MA, Gates N, Saigal N, Wilson GC, Meiklejohn J, Brodaty H, et al. The Study of Mental and Resistance Training (SMART) study—resistance training and/or cognitive training in mild cognitive impairment: a randomized, double-blind, double-sham controlled trial. *J Am Med Dir Assoc* 2014 Dec;15(12):873-880. [doi: [10.1016/j.jamda.2014.09.010](https://doi.org/10.1016/j.jamda.2014.09.010)] [Medline: [25444575](https://pubmed.ncbi.nlm.nih.gov/25444575/)]
23. Dang J, Zhang J, Guo Z, Lu W, Cai J, Shi Z, et al. A pilot study of iPad-assisted cognitive training for schizophrenia. *Arch Psychiatr Nurs* 2014 Jun;28(3):197-199. [doi: [10.1016/j.apnu.2014.01.003](https://doi.org/10.1016/j.apnu.2014.01.003)] [Medline: [24856273](https://pubmed.ncbi.nlm.nih.gov/24856273/)]
24. Des Roches CA, Balachandran I, Ascenso EM, Tripodis Y, Kiran S. Effectiveness of an impairment-based individualized rehabilitation program using an iPad-based software platform. *Front Hum Neurosci* 2014;8:1015 [FREE Full text] [doi: [10.3389/fnhum.2014.01015](https://doi.org/10.3389/fnhum.2014.01015)] [Medline: [25601831](https://pubmed.ncbi.nlm.nih.gov/25601831/)]
25. Godlove J, Anantha V, Advani M, Des Roches C, Kiran S. Comparison of therapy practice at home and in the clinic: a retrospective analysis of the constant therapy platform data set. *Front Neurol* 2019 Feb 25;10:140 [FREE Full text] [doi: [10.3389/fneur.2019.00140](https://doi.org/10.3389/fneur.2019.00140)] [Medline: [30858819](https://pubmed.ncbi.nlm.nih.gov/30858819/)]
26. Munsell M, De Oliveira E, Saxena S, Godlove J, Kiran S. Closing the digital divide in speech, language, and cognitive therapy: cohort study of the factors associated with technology usage for rehabilitation. *J Med Internet Res* 2020 Feb 07;22(2):e16286 [FREE Full text] [doi: [10.2196/16286](https://doi.org/10.2196/16286)] [Medline: [32044752](https://pubmed.ncbi.nlm.nih.gov/32044752/)]
27. Turunen M, Hokkanen L, Bäckman L, Stigsdotter-Neely A, Hänninen T, Paaanen T, et al. Computer-based cognitive training for older adults: determinants of adherence. *PLoS One* 2019 Jul 10;14(7):e0219541 [FREE Full text] [doi: [10.1371/journal.pone.0219541](https://doi.org/10.1371/journal.pone.0219541)] [Medline: [31291337](https://pubmed.ncbi.nlm.nih.gov/31291337/)]
28. 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](https://doi.org/10.1016/j.jalz.2011.03.005)] [Medline: [21514250](https://pubmed.ncbi.nlm.nih.gov/21514250/)]
29. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment 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):270-279 [FREE Full text] [doi: [10.1016/j.jalz.2011.03.008](https://doi.org/10.1016/j.jalz.2011.03.008)] [Medline: [21514249](https://pubmed.ncbi.nlm.nih.gov/21514249/)]
30. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005 Apr;53(4):695-699. [doi: [10.1111/j.1532-5415.2005.53221.x](https://doi.org/10.1111/j.1532-5415.2005.53221.x)] [Medline: [15817019](https://pubmed.ncbi.nlm.nih.gov/15817019/)]
31. Troyer AK, Rich JB. Psychometric properties of a new metamemory questionnaire for older adults. *J Gerontol B Psychol Sci Soc Sci* 2002 Jan;57(1):P19-P27. [doi: [10.1093/geronb/57.1.p19](https://doi.org/10.1093/geronb/57.1.p19)] [Medline: [11773220](https://pubmed.ncbi.nlm.nih.gov/11773220/)]
32. Weintraub S. The record of independent living: an informant-completed measure of activities of daily living and behavior in elderly patients with cognitive impairment. *Am J Alzheimer's Care Related Disorder* 1986 Apr 1;1(2):35-39. [doi: [10.1177/153331758600100210](https://doi.org/10.1177/153331758600100210)]
33. Logsdon RG, Gibbons LE, McCurry SM, Teri L. Assessing quality of life in older adults with cognitive impairment. *Psychosom Med* 2002;64(3):510-519. [doi: [10.1097/00006842-200205000-00016](https://doi.org/10.1097/00006842-200205000-00016)] [Medline: [12021425](https://pubmed.ncbi.nlm.nih.gov/12021425/)]
34. White T, Stern RA. Psychometric and technical manual. *Neuropsychological Assessment Battery*. 2003. URL: [https://www.parinc.com/WebUploads/productinfo/NAB/NAB\\_P&T\\_chap1.pdf](https://www.parinc.com/WebUploads/productinfo/NAB/NAB_P&T_chap1.pdf) [accessed 2022-06-29]
35. Zarit SH, Reever KE, Bach-Peterson J. Relatives of the impaired elderly: correlates of feelings of burden. *Gerontologist* 1980 Dec;20(6):649-655. [doi: [10.1093/geront/20.6.649](https://doi.org/10.1093/geront/20.6.649)] [Medline: [7203086](https://pubmed.ncbi.nlm.nih.gov/7203086/)]

36. Turk KW, Palumbo R, Deason RG, Marin A, Elshaar A, Gosselin E, et al. False memories: the other side of forgetting. *J Int Neuropsychol Soc* 2020 Jul;26(6):545-556 [FREE Full text] [doi: [10.1017/S1355617720000016](https://doi.org/10.1017/S1355617720000016)] [Medline: [32106906](https://pubmed.ncbi.nlm.nih.gov/32106906/)]
37. Constant Therapy- how it works. Constant Therapy Health. URL: <https://constanttherapyhealth.com/constant-therapy/> [accessed 2022-06-29]
38. Kiran S, Roches CD, Balachandran I, Ascenso E. Development of an impairment-based individualized treatment workflow using an iPad-based software platform. *Semin Speech Lang* 2014 Feb;35(1):38-50. [doi: [10.1055/s-0033-1362995](https://doi.org/10.1055/s-0033-1362995)] [Medline: [24449464](https://pubmed.ncbi.nlm.nih.gov/24449464/)]
39. 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](https://doi.org/10.1136/bmj.i5239)] [Medline: [27777223](https://pubmed.ncbi.nlm.nih.gov/27777223/)]
40. Kelley BJ, Petersen RC. Alzheimer's disease and mild cognitive impairment. *Neurol Clin* 2007 Aug;25(3):577-57v [FREE Full text] [doi: [10.1016/j.ncl.2007.03.008](https://doi.org/10.1016/j.ncl.2007.03.008)] [Medline: [17659182](https://pubmed.ncbi.nlm.nih.gov/17659182/)]
41. Galante E, Venturini G, Fiaccadori C. Computer-based cognitive intervention for dementia: preliminary results of a randomized clinical trial. *G Ital Med Lav Ergon* 2007;29(3 Suppl B):B26-B32. [Medline: [18575355](https://pubmed.ncbi.nlm.nih.gov/18575355/)]
42. Beck C, Heacock P, Mercer S, Thatcher R, Sparkman C. The impact of cognitive skills remediation training on persons with Alzheimer's disease or mixed dementia. *J Geriatr Psychiatry* 1988;21(1):73-88. [Medline: [3171111](https://pubmed.ncbi.nlm.nih.gov/3171111/)]
43. Herrera C, Chambon C, Michel BF, Paban V, Alescio-Lautier B. Positive effects of computer-based cognitive training in adults with mild cognitive impairment. *Neuropsychologia* 2012 Jul;50(8):1871-1881. [doi: [10.1016/j.neuropsychologia.2012.04.012](https://doi.org/10.1016/j.neuropsychologia.2012.04.012)] [Medline: [22525705](https://pubmed.ncbi.nlm.nih.gov/22525705/)]
44. Barnes DE, Yaffe K, Belfor N, Jagust WJ, DeCarli C, Reed BR, et al. Computer-based cognitive training for mild cognitive impairment: results from a pilot randomized, controlled trial. *Alzheimer Dis Assoc Disord* 2009;23(3):205-210 [FREE Full text] [doi: [10.1097/WAD.0b013e31819c6137](https://doi.org/10.1097/WAD.0b013e31819c6137)] [Medline: [19812460](https://pubmed.ncbi.nlm.nih.gov/19812460/)]
45. Yhnell E, Furby H, Lowe RS, Brookes-Howell LC, Drew CJ, Playle R, et al. A randomised feasibility study of computerised cognitive training as a therapeutic intervention for people with Huntington's disease (CogTrainHD). *Pilot Feasibility Stud* 2020;6:88 [FREE Full text] [doi: [10.1186/s40814-020-00623-z](https://doi.org/10.1186/s40814-020-00623-z)] [Medline: [32577299](https://pubmed.ncbi.nlm.nih.gov/32577299/)]
46. Tedim Cruz V, Pais J, Alves I, Ruano L, Mateus C, Barreto R, et al. Web-based cognitive training: patient adherence and intensity of treatment in an outpatient memory clinic. *J Med Internet Res* 2014 May 07;16(5):e122 [FREE Full text] [doi: [10.2196/jmir.3377](https://doi.org/10.2196/jmir.3377)] [Medline: [24808451](https://pubmed.ncbi.nlm.nih.gov/24808451/)]
47. Galvin JE, Tolea MI, Moore C, Chrisphonte S. The Number Symbol Coding Task: a brief measure of executive function to detect dementia and cognitive impairment. *PLoS One* 2020 Nov 30;15(11):e0242233 [FREE Full text] [doi: [10.1371/journal.pone.0242233](https://doi.org/10.1371/journal.pone.0242233)] [Medline: [33253192](https://pubmed.ncbi.nlm.nih.gov/33253192/)]
48. Mueller KD, Kosciak RL, LaRue A, Clark LR, Hermann B, Johnson SC, et al. Verbal fluency and early memory decline: results from the Wisconsin registry for Alzheimer's prevention. *Arch Clin Neuropsychol* 2015 Aug;30(5):448-457 [FREE Full text] [doi: [10.1093/arclin/acv030](https://doi.org/10.1093/arclin/acv030)] [Medline: [26025231](https://pubmed.ncbi.nlm.nih.gov/26025231/)]
49. Coyle H, Traynor V, Solowij N. Computerized and virtual reality cognitive training for individuals at high risk of cognitive decline: systematic review of the literature. *Am J Geriatr Psychiatry* 2015 Apr;23(4):335-359. [doi: [10.1016/j.jagp.2014.04.009](https://doi.org/10.1016/j.jagp.2014.04.009)] [Medline: [24998488](https://pubmed.ncbi.nlm.nih.gov/24998488/)]
50. Braley M, Pierce JS, Saxena S, De Oliveira E, Taraboanta L, Anantha V, et al. A virtual, randomized, control trial of a digital therapeutic for speech, language, and cognitive intervention in post-stroke persons with aphasia. *Front Neurol* 2021 Feb 12;12:626780 [FREE Full text] [doi: [10.3389/fneur.2021.626780](https://doi.org/10.3389/fneur.2021.626780)] [Medline: [33643204](https://pubmed.ncbi.nlm.nih.gov/33643204/)]
51. Edgar DL, Bargmann PL. Clinical study of the effectiveness of constant therapy in the treatment of clients with dementia: implications for telepractice. *Perspect ASHA SIGs* 2021 Jun 25;6(3):691-703. [doi: [10.1044/2021\\_persp-20-00123](https://doi.org/10.1044/2021_persp-20-00123)]
52. Bodner KA, Goldberg TE, Devanand DP, Doraiswamy PM. Advancing computerized cognitive training for MCI and Alzheimer's disease in a pandemic and post-pandemic world. *Front Psychiatry* 2020 Nov 25;11:557571 [FREE Full text] [doi: [10.3389/fpsyt.2020.557571](https://doi.org/10.3389/fpsyt.2020.557571)] [Medline: [33329097](https://pubmed.ncbi.nlm.nih.gov/33329097/)]
53. Hill NT, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. *Am J Psychiatry* 2017 Apr 01;174(4):329-340. [doi: [10.1176/appi.ajp.2016.16030360](https://doi.org/10.1176/appi.ajp.2016.16030360)] [Medline: [27838936](https://pubmed.ncbi.nlm.nih.gov/27838936/)]
54. Wu R, Rebok G. Maximizing the impact of cognitive engagement interventions for older adults. In: Thomas AK, Gutches A, editors. *The Cambridge Handbook of Cognitive Aging: A Life Course Perspective*. Cambridge, UK: Cambridge University Press; 2020:685-700.
55. Rogers W, Blocker K, Dupuy L. Current and emerging technologies for supporting successful aging. In: Thomas AK, Gutches A, editors. *The Cambridge Handbook of Cognitive Aging: A Life Course Perspective*. Cambridge, UK: Cambridge University Press; 2020:717-736.

## Abbreviations

**AD:** Alzheimer disease

**MCI:** mild cognitive impairment



**RBANS:** Repeatable Battery for the Assessment of Neuropsychological Status

**RCT:** randomized controlled trial

**TBI:** traumatic brain injury

*Edited by A Mavragani; submitted 23.10.21; peer-reviewed by J Perales, D Mohan; comments to author 25.03.22; revised version received 30.06.22; accepted 11.07.22; published 12.09.22.*

*Please cite as:*

Marin A, DeCaro R, Schiloski K, Elshaar A, Dwyer B, Vives-Rodriguez A, Palumbo R, Turk K, Budson A  
*Home-Based Electronic Cognitive Therapy in Patients With Alzheimer Disease: Feasibility Randomized Controlled Trial*  
*JMIR Form Res* 2022;6(9):e34450

URL: <https://formative.jmir.org/2022/9/e34450>

doi: [10.2196/34450](https://doi.org/10.2196/34450)

PMID: [36094804](https://pubmed.ncbi.nlm.nih.gov/36094804/)

©Anna Marin, Renée DeCaro, Kylie Schiloski, Ala'a Elshaar, Brigid Dwyer, Ana Vives-Rodriguez, Rocco Palumbo, Katherine Turk, Andrew Budson. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 12.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Predicting Depression in Patients With Knee Osteoarthritis Using Machine Learning: Model Development and Validation Study

Zuzanna Nowinka<sup>1</sup>, BSc, MBChB; M Abdulhadi Alagha<sup>1,2</sup>, MD; Khadija Mahmoud<sup>1</sup>, BSc; Gareth G Jones<sup>1</sup>, PhD

<sup>1</sup>MSk Lab, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, United Kingdom

<sup>2</sup>Data Science Institute, London School of Economics and Political Science, London, United Kingdom

**Corresponding Author:**

M Abdulhadi Alagha, MD

MSk Lab

Department of Surgery and Cancer, Faculty of Medicine

Imperial College London

South Kensington Campus

London, SW7 2AZ

United Kingdom

Phone: 44 020 7589 5111

Email: [h.alagha@imperial.ac.uk](mailto:h.alagha@imperial.ac.uk)

## Abstract

**Background:** Knee osteoarthritis (OA) is the most common form of OA and a leading cause of disability worldwide. Chronic pain and functional loss secondary to knee OA put patients at risk of developing depression, which can also impair their treatment response. However, no tools exist to assist clinicians in identifying patients at risk. Machine learning (ML) predictive models may offer a solution. We investigated whether ML models could predict the development of depression in patients with knee OA and examined which features are the most predictive.

**Objective:** The primary aim of this study was to develop and test an ML model to predict depression in patients with knee OA at 2 years and to validate the models using an external data set. The secondary aim was to identify the most important predictive features used by the ML algorithms.

**Methods:** Osteoarthritis Initiative Study (OAI) data were used for model development and external validation was performed using Multicenter Osteoarthritis Study (MOST) data. Forty-two features were selected, which denoted routinely collected demographic and clinical data such as patient demographics, past medical history, knee OA history, baseline examination findings, and patient-reported outcome measures. Six different ML classification models were trained (logistic regression, least absolute shrinkage and selection operator [LASSO], ridge regression, decision tree, random forest, and gradient boosting machine). The primary outcome was to predict depression at 2 years following study enrollment. The presence of depression was defined using the Center for Epidemiological Studies Depression Scale. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC) and F1 score. The most important features were extracted from the best-performing model on external validation.

**Results:** A total of 5947 patients were included in this study, with 2969 in the training set, 742 in the test set, and 2236 in the external validation set. For the test set, the AUC ranged from 0.673 (95% CI 0.604-0.742) to 0.869 (95% CI 0.824-0.913), with an F1 score of 0.435 to 0.490. On external validation, the AUC varied from 0.720 (95% CI 0.685-0.755) to 0.876 (95% CI 0.853-0.899), with an F1 score of 0.456 to 0.563. LASSO modeling offered the highest predictive performance. Blood pressure, baseline depression score, knee pain and stiffness, and quality of life were the most predictive features.

**Conclusions:** To our knowledge, this is the first study to apply ML classification models to predict depression in patients with knee OA. Our study showed that ML models can deliver a clinically acceptable level of performance (AUC>0.7) in predicting the development of depression using routinely available demographic and clinical data. Further work is required to address the class imbalance in the training data and to evaluate the clinical utility of the models in facilitating early intervention and improved outcomes.

(JMIR Form Res 2022;6(9):e36130) doi:[10.2196/36130](https://doi.org/10.2196/36130)

**KEYWORDS**

knee osteoarthritis; depression; machine learning; predictive modeling

## Introduction

Knee osteoarthritis (OA) is the most common form of OA and a leading cause of disability worldwide, with global prevalence estimated at 16% for individuals aged 15 years and over [1]. Knee OA is a chronic, progressive condition characterized by structural damage to the cartilage [2]. Knee OA results in chronic pain and impaired joint function, significantly limiting the activities of daily living [1,3]. Consequently, these patients experience a poorer health-related quality of life and are at higher risk of developing depression compared to the general population [4]. It has been estimated that up to 20% of patients with knee OA may be suffering from depression [3].

Several studies suggest that depression has an adverse impact on OA prognosis, quality of life, pain levels, as well as treatment effectiveness [5-7]. A longitudinal study conducted by Rathbun et al [8] found that depressive symptoms affected the physical functioning and pain severity of patients with knee OA. Another study showed that a persistently depressed mood significantly increases the severity of pain [9]. Additionally, a bidirectional relationship between pain and depression in patients with knee OA has been described, where concurrent depression increases pain perception and, reciprocally, higher pain levels may lead to a more depressed state [9-11]. It is therefore essential to recognize and address the vicious pain-depression cycle early.

Unsurprisingly, patients with knee OA and comorbid depression report lower coping ability, which translates into more frequent medical help-seeking and reduced satisfaction from treatment, including surgical interventions such as knee arthroplasty [3,10,12,13]. Ultimately, this accounts for a substantial rise in the health care cost burden [14,15]. Agarwal et al [16] estimated that the health care costs per year increase by US \$4400 (US \$13,684 vs US \$9284) for every patient with concurrent OA and depression. The economic cost associated with knee OA is likely to rise in the upcoming years due to increasing life expectancy and thus the proportion of patients with knee OA [2]. With no curative treatment in sight, emphasis should be made on preventative and nonoperative strategies to manage the disease symptoms and reduce worsening factors such as depression [1,12].

Obtaining adequate mental health support should be of primary importance, as the presence of depressive symptoms is a significant predictor of worsening outcomes [17]. At the same time, appropriate therapy with antidepressants and counseling has been shown to significantly lower the perceived severity of pain [18]. However, less than half of all patients affected by knee OA and concurrent depression actively seek support or receive adequate treatment [19,20]. Unfortunately, poor mental health is frequently overlooked by clinicians, who focus primarily on the physical aspects of knee OA and so fail to recognize depression or its role in contributing to persisting knee symptoms [12,21]. Being able to predict which patients are at risk of experiencing depression would facilitate a targeted,

preventative strategy against worsening outcomes such as pain and declining physical function [17].

Identifying patients with depression early would be helpful; however, no such tools currently exist. Although one previous study has tried to predict depression in this patient population, the model was based on conventional statistical methods, had low accuracy (area under the receiver operating characteristic curve [AUC]=0.742, 95% CI 0.622-0.862), and lacked external validation [22]. This represents a significant gap in care. The solution may lie in machine learning (ML) models. The ability of ML algorithms to handle large data sets, and evaluate complex and nonlinear relationships between variables theoretically makes them better suited for predictive tasks than standard statistical methods [23,24]. To date, no previous study has attempted to build an ML prediction model to detect the development of depression in patients with knee OA.

The primary objective of this study was to apply ML models to predict depression in patients with knee OA, using routinely available clinical data. We hypothesized that ML models can deliver a clinically acceptable level of performance, defined as an AUC greater than 0.7. Our secondary objective was to identify the most important predictive features used by the ML algorithms to make this prediction.

## Methods

### Data Sources and Study Cohort

We used data from the Osteoarthritis Initiative (OAI) database for model development and data from the Multicenter Osteoarthritis Study (MOST) for external validation. Both are publicly available, prospective cohort studies investigating knee OA progression in the US population [25,26]. The OAI study included adults aged 45-79 years, enrolled between February 2004 and May 2006, and the MOST included adults aged 50-79 years, recruited in 2003.

We included patients who attended the baseline and 15-month/24-month follow-ups, with preexisting knee OA (defined as the presence of symptoms and radiographic evidence of OA) or at high risk of developing knee OA (symptoms of pain, stiffness, and swelling). Patients with a history of rheumatoid arthritis, missing data for the depression scale scores at either consultation, missing radiographic data, missing baseline examination findings, or missing patient-reported outcome measures were excluded.

### Ethics Considerations

No ethical approval was required for this study owing to the open access nature of the OAI and MOST databases.

### Prediction Outcome

Our primary outcome was the development of depression at 2 years following enrollment in the database. Depression was defined using the Center for Epidemiological Studies Depression Scale (CES-D), which is based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition formulation of

depression, containing 20 questions evaluating the severity of psychosomatic symptoms [27]. The score ranges from 0 to 60, with higher values indicating greater symptom severity. A score of 16 points or more has previously been linked to clinical depression and as such was used in this study to dichotomize patients as either depressed or not depressed [27].

In the MOST, follow-up visits were scheduled at different time points compared with those used in the OAI study, and therefore CES-D scores captured during the 15-month visit were used for external validation.

### Variable Selection

Variable selection was guided by the literature and clinical relevance as judged by the senior author who is a specialist in the field. To facilitate external validation, equivalent variables had to be available in both the OAI and MOST data sets. In total, there were 2532 baseline variables in the OAI database and 1842 baseline variables in the MOST database; 70 and 66 variables were selected from the respective databases for model development. Variables included information on patient demographics, past medical history, knee OA history, baseline examination findings, and baseline patient-reported outcome measures.

Patient demographics included age, sex, ethnicity, BMI, marital status, living arrangements, current employment, education, and smoking status. Past medical history encompassed the history of heart attack, heart failure, stroke, asthma, chronic obstructive pulmonary disease, peptic ulcer disease, diabetes, kidney disease, and osteoporosis medication. Variables relating to knee OA history consisted of past knee injury, past knee surgery, steroid knee injections, analgesic medication for knee pain, as well as other arthritis medication. Baseline examination findings covered systolic and diastolic blood pressure, medial and lateral tibiofemoral, Kellgren-Lawrence grade, the 20-meter-walk test, the five-times-sit-to-stand test, and baseline CES-D score. Patient-reported outcome measures were the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Physical Activity Scale for the Elderly (PASE), and 12-item Short-Form Health Survey (SF-12).

## Data Preprocessing

### Binning the Features

Smoking status was stratified according to smoking intensity into light (1-5 pack-year history of smoking), moderate (10-20 pack-years), or severe (>20 pack-years). BMI was grouped into underweight (BMI<18.5 kg/m<sup>2</sup>), normal weight (BMI 18.5-24.9 kg/m<sup>2</sup>), overweight (BMI 25-29.9 kg/m<sup>2</sup>), and obese (BMI>30 kg/m<sup>2</sup>), as defined by the World Health Organization [28]. Patients were categorized according to the American Heart Association Hypertension Guidelines to denote the stage of hypertension using variables for systolic and diastolic blood pressures [29]. Results of the five-times-sit-to-stand test were dichotomized, given that ≥10 seconds is the optimal cutoff for predicting the development of disability [30].

### Feature Engineering

Feature engineering involves the combination of separate variables into a new, “engineered” feature, based on domain expertise and literature evidence. This action decreases the number of separate features and has been shown to improve model performance [31]. The “ethnicity” feature was created by merging variables describing race (*white, Black, Hispanic, other*). Variables assessing living arrangements were combined to denote whether the patient lived alone or with someone else. A feature for OA history was created by combining variables denoting the presence of other types of arthritis (*no other arthritis, one or more joints affected by OA, gout, OA and gout*). Variables denoting the use of analgesic medication for knee OA were assigned into a single feature, “analgesic medication” (*no pain relief, topical salicylates, nonsteroidal anti-inflammatory drugs or cyclooxygenase-2 inhibitors, opioid medication, combination of analgesic medication, other*). The “OA medication” feature was created by combining variables with information on OA treatment and supplements (*no medication or vitamin D supplements, bisphosphonates, estrogen/raloxifene, calcitonin/teriparatide, combination of OA medications*). The “arthritis medication” feature was created by merging five variables (*oral corticosteroids, supplements*). The final list of 42 features included in model training is summarized in [Table 1](#).

**Table 1.** Summary of all features included in the model training.

Feature category	Features
Patient demographics	Age, sex, BMI, ethnicity, employment status, education status, living alone, marital status, smoking status
Past medical history and medication	Heart attack, heart failure, stroke, asthma, chronic obstructive pulmonary disease, peptic ulcer disease, diabetes, kidney disease, osteoporosis medication
Knee osteoarthritis history	Knee arthroscopy, knee meniscectomy, ligament repair, other knee surgery, arthritis of other joints, knee injury, steroid knee injections, analgesic medication for knee osteoarthritis, arthritis medication
Baseline examination findings	Blood pressure, 20-meter-walk test, five-stands-to-sit test, KLG <sup>a,b</sup> , CES-D <sup>c</sup> baseline
Patient-reported outcome measures	WOMAC <sup>a,d</sup> (Total, Pain score, Stiffness score); SF-12 <sup>e</sup> (Physical components, Mental health component); PASE <sup>f</sup>

<sup>a</sup>Separate feature for the right and left knee.

<sup>b</sup>KLG: Kellgren-Lawrence Grade.

<sup>c</sup>CES-D: Center for Epidemiological Studies Depression Scale.

<sup>d</sup>WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

<sup>e</sup>SF-12: 12-item Short Form Health Survey.

<sup>f</sup>PASE: Physical Activity Scale for the Elderly.

### Missing Values

Missing values in the OAI data set were addressed by coding them as “unknown” to match the MOST data set. Following this imputation, only patients with all observations completed were included for analysis.

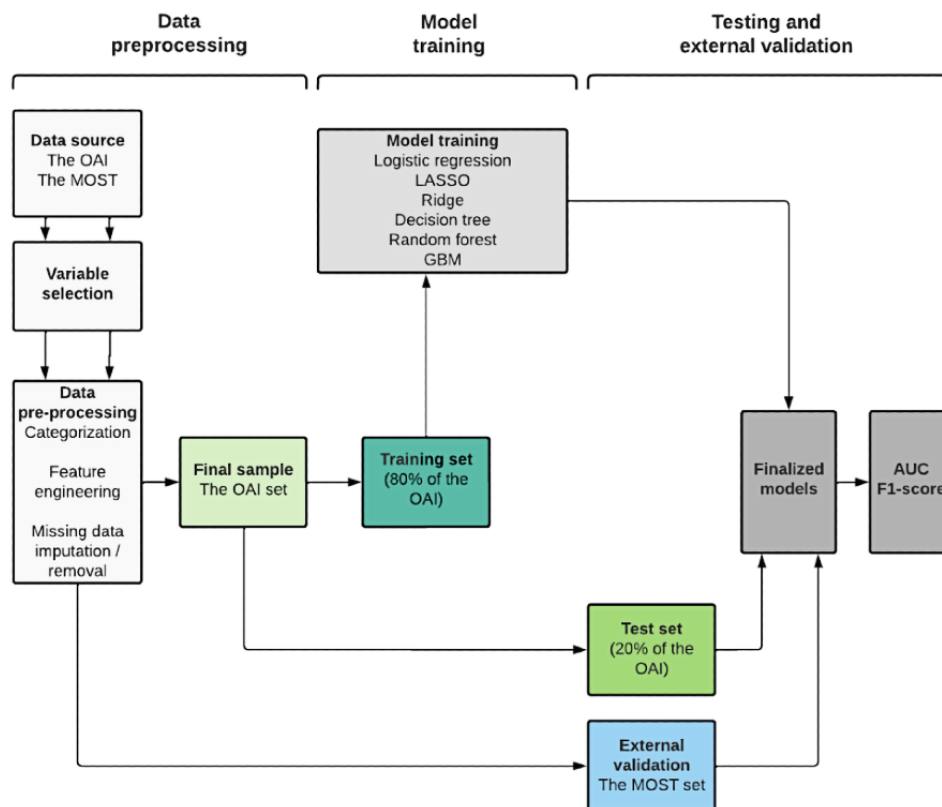
### Model Development

#### Overview

Figure 1 summarizes the stages of data preprocessing and model development. The OAI data set was randomly divided into training (80% of observations) and test (20% of observations) sets using a computer algorithm, ensuring that each set included an equal proportion of patients with depression. Six common

classification ML algorithms (logistic regression, least absolute shrinkage and selection operator [LASSO], ridge, decision tree, random forest, and gradient boosting machine [GBM]) were trained using the same set of 42 features. Classification models are a type of supervised ML where the algorithm calculates a probability of an observation belonging to the “positive” class based on the input data [32]. If the probability is above the threshold, the observation is labeled as “positive” (ie, depressed). The probability threshold is by default set to 0.5 but can be lowered when the cost of missing a “positive” case is high. Therefore, in this study, the threshold was set to 0.2 [33]. For each model, hyperparameter tuning was conducted until the performance on the training set was maximized. All models were developed using RStudio software (version 1.4.1106) [34].

**Figure 1.** Flowchart summarizing the project timeline and steps of model development. AUC: area under the receiver operating characteristic curve; GBM: gradient boosting machine; LASSO: least absolute shrinkage and selection operator; MOST: Multicenter Osteoarthritis Study; OAI: Osteoarthritis Initiative.



### Logistic Regression

Logistic regression is a statistical model that uses a logit function to predict the probability of an observation belonging to the positive class [35]. Logistic regression is well-suited for classification problems such as problems involving describing the risk of developing a disease or the risk of mortality. This model was implemented using the RStudio “stats” package [36].

### LASSO and Ridge Regression

LASSO and ridge regression models are based on the logistic regression model [24,32,37]. In LASSO, the algorithm adds a “penalty” to each feature so that features are eliminated if not considered important for the prediction by the algorithm [37]. LASSO shrinks regression coefficients toward 0, and ultimately only top informative features are included. This results in a simpler and more easily interpretable model [37]. In ridge, the algorithm reduces less important features to close to zero but does not eliminate them [32]. In this way, all features are kept in the model, which is beneficial when all features need to be included [32]. LASSO and ridge models were developed using the “glmnet” package with optimal hyperparameters for both algorithms set as follows:  $n\text{folds}=3$ ,  $s=\text{lambda.min}$  [38].

### Decision Tree and Random Forest

Decision tree is a simple, tree-shaped algorithm, in which each branch of the tree determines a possible decision or course of action [39]. The model was developed with no additional hyperparameters using the “rpart” package [40]. Random forest is an algorithm similar to the decision tree; it operates by building multiple, independently trained decision trees using

random subsets of the data [41]. Subsequently, their predictions are combined into a single prediction outcome. Random forest of 500 trees with  $\text{nodesize}=100$  and  $\text{mtry}=4$  was developed using the “randomForest” package [42].

### GBM Model

In GBM, multiple tree-based classifiers are trained to augment each other and to reduce the prediction error [43]. GBM differs from the random forest algorithm in that a new decision tree is trained with the aim to correct errors made by existing trees, rather than training them independently. This model was developed using the “gbm” package and optimum hyperparameters were  $\text{ntrees}=2000$ ,  $\text{cv.folds}=3$ ,  $\text{interaction.depth}=4$ , and  $\text{shrinkage}=0.1$  [44].

### Performance Evaluation

The overall model performance was evaluated on the previously unseen OAI test set and externally validated using the MOST data set.

The primary model performance criterion was the AUC, and we considered an AUC greater than 0.7 to indicate clinically acceptable performance [45]. For each model, accuracy, precision, and recall are also reported. In addition, the F1 score, a weighed metric of precision and recall, was calculated according to the formula:  $F1=2 \times ([\text{precision} \times \text{recall}] / [\text{precision} + \text{recall}])$ . F1 score ranges from 0 (poor performance) to 1 (perfect performance).

While ML may provide a valuable predictive tool, the clinical implementation often raises concerns due to the model’s complexity, referred to as the “black-box” problem [46]. One

way of improving model understanding is by extracting the most important predictive features. We therefore identified the most important predictive features from the best-performing model.

## Results

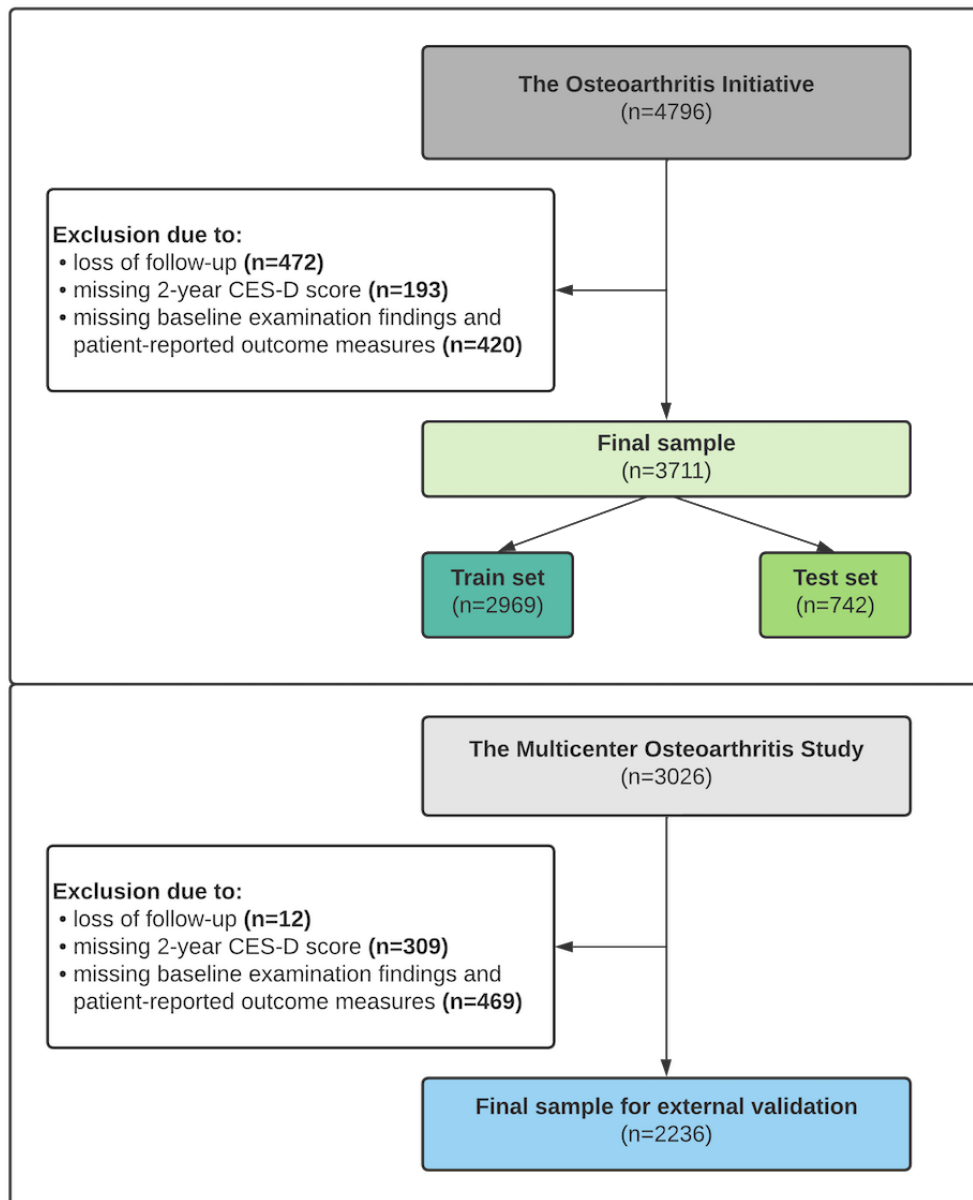
### Study Participants

The initial OAI data set included 4796 patients (Figure 2). Following exclusion of 1085 patients, the final sample size encompassed 3711 patients. After splitting the sample, the training set included 2969 patients and the test set had 742 observations. In the MOST data set, 790 patients were excluded

from the initial sample of 3026 cases and the final sample included 2236 patients.

Table 2 summarizes the key patient characteristics. The average age was 61.0 years for the OAI sample and 62.1 years for the MOST sample. In both data sets, the majority of patients were female and of white ethnicity. Less than half of the patients had hypertension stage 1 or higher. There were some differences between the OAI and MOST samples. First, the proportion of depressed patients at 2 years was higher in the MOST sample. The MOST population also had higher average WOMAC scores for both the right and left knees, and a greater proportion of patients using analgesic medication for knee OA.

Figure 2. Summary of patient flow for both databases. CES-D: Center for Epidemiological Studies Depression Scale.



**Table 2.** Key patient demographic and clinical data.

Characteristic	OAI <sup>a</sup> (n=3711)	MOST <sup>b</sup> (n=2236)
Age, mean (SD)	61.0 (9.1)	62.1 (8.1)
BMI, mean (SD)	28.4 (4.8)	30.4 (5.9)
Sex (female), n (%)	2149 (57.91)	1297 (58.01)
Ethnicity (white), n (%)	3082 (83.05)	1932 (86.40)
Blood pressure (hypertension stage $\geq$ 1), n (%)	1847 (49.77)	1008 (45.08)
Other arthritis, n (%)	1454 (39.18)	1071 (47.90)
Analgesic medication for knee OA <sup>c</sup> (any), n (%)	845 (22.77)	1804 (80.68)
<b>KLG<sup>d</sup>, n (%)</b>		
Right knee, grade 1 or higher	2294 (61.82)	1180 (52.77)
Left knee, grade 1 or higher	2206 (59.44)	1264 (56.53)
<b>WOMAC<sup>e</sup>-total, mean (SD)</b>		
Right knee	10.7 (10.3)	18.6 (17.5)
Left knee	10.7 (10.4)	18.3 (17.5)
Baseline CES-D <sup>f</sup> , mean (SD)	6.3 (6.0)	6.7 (6.2)
Depression at 2-year visit, n (%)	342 (9.22)	265 (11.85)

<sup>a</sup>OAI: Osteoarthritis Initiative.

<sup>b</sup>MOST: Multicenter Osteoarthritis Study.

<sup>c</sup>OA: osteoarthritis.

<sup>d</sup>KLG: Kellgren-Lawrence Grade.

<sup>e</sup>WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

<sup>f</sup>CES-D: Center for Epidemiological Studies Depression Scale.

## Model Performance

In total, six classification models were trained using all 42 features. The results for each model are summarized in [Table 3](#). [Figure 3](#) and [Figure 4](#) present the AUC plots for the internal test set and the external validation set, respectively. The AUC ranged from 0.673 to 0.869 for the internal test set and from 0.720 to 0.876 for the external validation set. Except for the decision tree algorithm, all models yielded an AUC $>$ 0.7, suggesting clinically acceptable discrimination between depressed and nondepressed patients [45]. LASSO was the model with the highest AUC on both the internal test set and external validation set.

The accuracy, precision, recall, and F1 scores for the test and validation sets are summarized in [Table 4](#) and [Table 5](#), respectively. The accuracy on the OAI test set varied from 0.895 (decision tree) to 0.923 (random forest). The performance on this metric was lower for the MOST data set, ranging from 0.865 (GBM) to 0.895 (ridge). Despite high accuracy, the proportion of correctly classified positive cases was relatively low. For the internal test set, the F1 scores varied from 0.435 (decision tree) to 0.490 (LASSO), and from 0.456 (ridge) to 0.536 (LASSO) on external validation. LASSO had a consistently high performance for the AUC and F1 score in comparison to the other models, ranking first on both the internal test and external validation sets.



**Table 3.** Model performance for the internal test set and external validation set.

Rank <sup>a</sup>	Model	Test set (OAI <sup>b</sup> ), AUC <sup>c</sup> (95% CI)	External validation set (MOST <sup>d</sup> ), AUC (95% CI)
1	LASSO <sup>e</sup>	0.869 (0.824-0.913)	0.876 (0.853-0.899)
2	GBM <sup>f</sup>	0.858 (0.813-0.903)	0.872 (0.849-0.895)
3	Ridge	0.864 (0.818-0.910)	0.852 (0.827-0.878)
4	Random forest	0.808 (0.741-0.874)	0.822 (0.790-0.853)
5	Logistic regression	0.837 (0.786-0.888)	0.808 (0.775-0.840)
6	Decision tree	0.673 (0.604-0.742)	0.720 (0.685-0.755)

<sup>a</sup>Models are ranked by their performance on the external validation data set.

<sup>b</sup>OAI: Osteoarthritis Initiative.

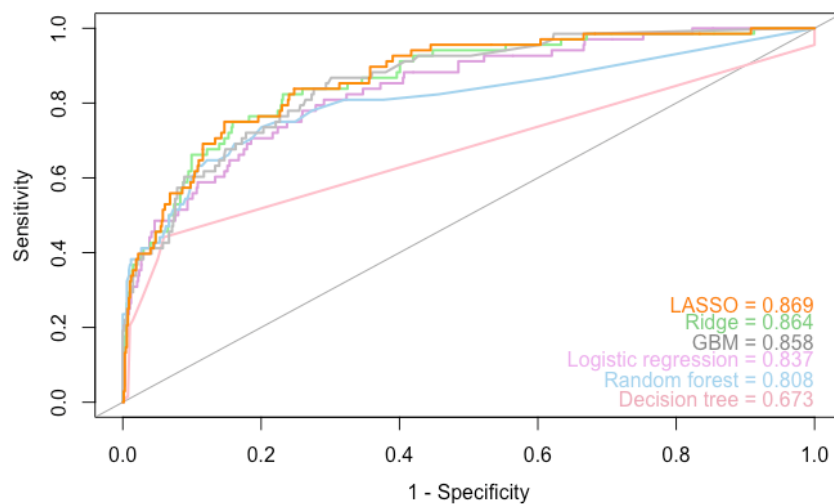
<sup>c</sup>AUC: area under the receiver operating characteristic curve.

<sup>d</sup>MOST: Multicenter Osteoarthritis Study.

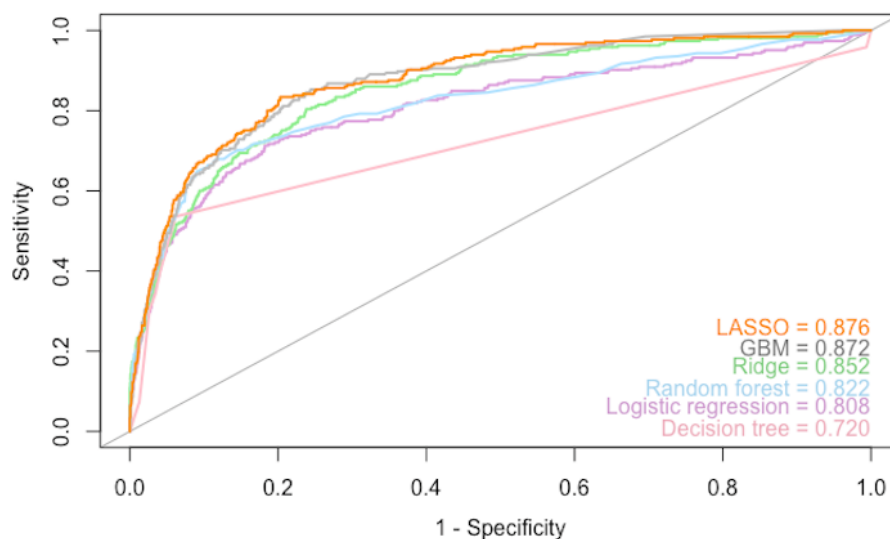
<sup>e</sup>LASSO: least absolute shrinkage and selection operator.

<sup>f</sup>GBM: gradient boosting machine.

**Figure 3.** AUC plot of all models tested on the OAI test set (20% of the initial OAI data set). The test set was not used at any stage of model training. AUC: area under the receiver operating characteristic curve; GBM: gradient boosting machine; LASSO: least absolute shrinkage and selection operator; MOST: Multicenter Osteoarthritis Study; OAI: Osteoarthritis Initiative.



**Figure 4.** AUC plot of all models externally validated on the MOST data set. AUC: area under the receiver operating characteristic curve; GBM: gradient boosting machine; LASSO: least absolute shrinkage and selection operator; MOST: Multicenter Osteoarthritis Study; OAI: Osteoarthritis Initiative.



**Table 4.** Accuracy, precision, recall, and F1 scores for the test set, ranked by the F1 score.

Rank	Model	Accuracy	Precision	Recall	F1
1	LASSO <sup>a</sup>	0.902	0.467	0.515	0.490
2	Random forest	0.923	0.628	0.397	0.486
3	Logistic regression	0.906	0.485	0.485	0.485
4	GBM <sup>b</sup>	0.901	0.466	0.500	0.482
5	Decision tree	0.895	0.429	0.441	0.435
6	Ridge	0.908	0.500	0.426	0.460

<sup>a</sup>LASSO: least absolute shrinkage and selection operator.

<sup>b</sup>GBM: gradient boosting machine.

**Table 5.** Accuracy, precision, recall, and F1 scores for the validation set, ranked by the F1 score.

Rank	Model	Accuracy	Precision	Recall	F1
1	LASSO <sup>a</sup>	0.889	0.528	0.604	0.563
2	Decision tree	0.890	0.538	0.536	0.537
3	GBM <sup>b</sup>	0.865	0.453	0.657	0.536
4	Random forest	0.894	0.556	0.506	0.530
5	Logistic regression	0.886	0.344	0.698	0.461
6	Ridge	0.895	0.593	0.370	0.456

<sup>a</sup>LASSO: least absolute shrinkage and selection operator.

<sup>b</sup>GBM: gradient boosting machine.

## Most Important Predictive Features

The most important predictive features identified by LASSO were blood pressure, CES-D score at baseline, total WOMAC score for both knees, and mental and physical components of the SF-12 survey. Blood pressure had the highest coefficient (0.173), followed by the baseline CES-D score (0.126), WOMAC total for the right knee (0.004), and WOMAC total for the left knee (0.003). The mental and physical components of SF-12 had negative coefficients (−0.032 and −0.009, respectively).

## Discussion

### Principal Findings

The results of this study demonstrate that it is possible, with high accuracy, to predict depression in patients with knee OA using a variety of routinely collected data such as patient demographics, medical history, examination findings, and patient-reported outcome measures. The developed ML models achieved clinically relevant discrimination between depressed and nondepressed patients, with LASSO identified as the best-performing model, yielding an AUC of 0.876 (95% CI 0.853–0.899) on external validation. The accuracies for external validation were high, ranging from 0.865 (GBM) to 0.895 (ridge), meaning that between 86.5% and 89.5% of all patients were correctly classified. However, the F1 scores ranged from 0.456 (ridge) to 0.563 (LASSO). Low F1 scores despite high accuracy implies that the models can identify patients without depression more accurately than those with depression. This is

likely due to class imbalance in the data set, which is a common problem in medical research that results in predictive modeling bias toward the majority [47].

While ML may provide a valuable predictive tool, the clinical implementation often raises concerns due to model complexity, often referred to as the “black-box” problem [46]. One way of improving model understanding is to extract the most important features [48]. In this study, blood pressure, the baseline CES-D, the total WOMAC, as well as mental and physical components for SF-12 were identified as being the most informative measures for prediction. Although this does not imply a statistically significant correlation between the features and the prediction outcome, it is reassuring that the input features identified by LASSO have previously been highlighted as factors associated with an increased risk of developing depression in patients with OA [8,9,49]. Surprisingly, blood pressure was identified as being the most informative factor for prediction. The presence of multiple comorbidities can further increase the risk of depression development in patients with knee OA, regardless of their pathophysiology [49]. Notably, the radiographic severity of OA was not highlighted as a predictive feature for depression development. This is consistent with previous research showing that depression and pain are independent from the extent of radiographic degenerative changes [50]. This known discrepancy between knee OA symptoms and radiographic severity highlights the complex nature of the disease and the need for more objective assessment tools. The association between depression, chronic conditions, and pain is complex. The temporality of the relationship between depression and pain has been poorly researched, but it appears

that both factors potentiate each other, with higher pain severity increasing the persistence of depressed mood and the presence of pain increasing the incidence of depression [5,7,28,51,52]. This highlights the essential role of appropriate, interdisciplinary mental health support for patients with knee OA.

ML predictive models have an important role in augmenting clinical judgment, and when compared with standard predictions, they produce more accurate and less variable risk estimates [53]. The best-performing model in our study, LASSO, could be potentially used to aid in identification of patients at risk of future depression. Since the CES-D score has been designed as a screening tool, the patients identified as “positive” by our model would have to undergo further, more specialist mental health assessment. Depending on that outcome, the patients could be offered either a self-help aid, or potentially, a specialist referral. This would be more economical and time-efficient than assessing every patient attending with knee pain. However, further research is required since the implementation of predictive models is often difficult due to lack of clear clinical guidance on how to act upon the predicted outcome [54].

The advantage of our models lies in their simplicity as they rely on easily accessible clinical information. In addition, LASSO identified only 6 features to be crucial for prediction, making the model more practical. Blood pressure is routinely measured by primary health care practitioners, and WOMAC, SF-12, and CES-D scores are commonly used patient-reported outcome measures [55-57]. The aforementioned questionnaires are brief and require minimal training. Currently, there is no proven strategy to prevent or cure knee OA, and the therapy is focused on alleviating pain and addressing functional limitations [9]. Since depression is a potentially modifiable risk factor for worsening pain and function in knee OA, our prediction model could offer a targeted, preventative strategy. Diagnosing depression in patients with concurrent chronic pain conditions is challenging and having such information would facilitate discussions around the patient’s mental health, even at times when the patient is not yet aware of their symptoms. While further research is required to evaluate the practical aspects of the clinical application, the findings of our study represent an important step toward developing a potential diagnostic aid, addressing a significant gap in knee OA care.

### Comparison With Prior Work

To the best of our knowledge, this is the first study applying ML to predict depression in patients with knee OA. One previous study attempted to develop a prediction model based on logistic regression using conventional statistical methods [22]. Although the model achieved a clinically acceptable performance with an AUC of 0.742 (95% CI 0.622-0.862), it was built using a small sample of patients and was not tested on an independent sample or externally validated [22].

Diagnosis of depression is challenging in clinical practice, and ML models have been previously applied to predict illness in different patient populations [58-62]. Clinically relevant predictive performance of common ML classification algorithms was shown in two studies predicting postpartum depression [58,59]. Cvetkovic [60] used a deep-learning approach to predict depression in breast cancer patients, achieving high internal

accuracy. However, the study methodology was poorly reported, with information lacking on data preprocessing and model testing [60]. In another study, depression and anxiety in college students were estimated using GBM, with satisfactory performance yielding an AUC of 0.730 [61]. When applied to community-residing older adults, a logistic regression model achieved variable accuracy, ranging from 58.33% for severe depression to 90.44% for mild depression [62]. The variation in model performance achieved by these studies could be attributed to the use of different algorithms, different evaluation tools for detection of depressive symptoms, as well as the use of different predictive features.

### Strengths

Our study is strengthened by the use of a large patient cohort for model development, testing, and validation. The list of input features was carefully curated, with selection based on literature evidence, domain expertise, and data completeness. In addition, our predictive models were externally validated and performed well in an independent cohort, demonstrating their generalizability and potential for clinical application. Notably, LASSO identified only six features to be crucial for prediction, which showcases the simplicity of our method and the ease with which this tool could be used in a clinical setting.

### Limitations

Several limitations should be addressed in future research. First, the study sample used for model development might not be representative of a general population of patients with knee OA. The prevalence of depressed patients in the training set was 9.2%, which is much lower than the 20% rate previously suggested by the literature [63]. The OAI study excluded patients with end-stage OA, morbid obesity, or those with terminal diseases, whereas these factors are associated with an even higher risk of depression [25,49]. Second, both the OAI and the MOST data sets were based in the United States with patients from a predominantly white ethnic background [25,26]. Further validation of our prediction model in a more ethnically and socioeconomically diverse population would help to detect any potential discrimination. Third, due to differences in the OAI and MOST protocols, follow-up times differed by 15 months between the training and external validation sets. Nevertheless, the models were able to predict on the external data set with similar performance. Lastly, the presence of depression at 2 years was defined using the CES-D scale; although this tool has been validated for use in patients with chronic illness and OA, it is not considered a gold standard for the diagnosis of depression [27]. However, the CES-D questionnaire has the advantage of being brief, easy to understand, and requiring minimal training for the assessor [27].

### Conclusions

This is the first study to apply ML classification models to predict depression in patients with knee OA using routinely collected patient data. The LASSO model offered the highest quality of prediction, with an AUC of 0.876 (95% CI 0.853-0.899) on external validation. The advantages of our method include the use of a large patient cohort and routinely collected data, as well as external validation on an independent

data set. This tool offers a potential opportunity to assess a patient's risk of future depression, facilitating early intervention. Further research is required to establish where such a tool would fit within the care pathway, and while the harmful effects of depression on knee OA are well documented, it will be necessary to confirm that early detection and management of depression in this population leads to the expected improvement in outcomes.

## Acknowledgments

MAA is funded by the Imperial College President's PhD Scholarship.

## Authors' Contributions

ZN, MAA, KM, and GGJ were involved in setting out the project aim and methodology. ZN conducted the literature search and wrote the original draft. ZN, MAA, and KM contributed to data curation and analysis. MAA and GGJ contributed to study design. GGJ supervised the conduction of the study, and reviewed and edited the manuscript. All authors had access to the raw data and have approved the final manuscript.

## Conflicts of Interest

None declared.

## References

1. Cui A, Li H, Wang D, Zhong J, Chen Y, Lu H. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. *EClinicalMedicine* 2020 Dec;29-30:100587 [FREE Full text] [doi: [10.1016/j.eclinm.2020.100587](https://doi.org/10.1016/j.eclinm.2020.100587)] [Medline: [34505846](https://pubmed.ncbi.nlm.nih.gov/34505846/)]
2. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *The Lancet* 2019 Apr 27;393(10182):1745-1759. [doi: [10.1016/S0140-6736\(19\)30417-9](https://doi.org/10.1016/S0140-6736(19)30417-9)] [Medline: [31034380](https://pubmed.ncbi.nlm.nih.gov/31034380/)]
3. Marks R. Depression and osteoarthritis: impact on disability. *Aging Sci* 2014;02(03):1000126. [doi: [10.4172/2329-8847.1000126](https://doi.org/10.4172/2329-8847.1000126)]
4. van 't Land H, Verdurmen J, Ten Have M, van Dorsselaer S, Beekman A, de Graaf R. The association between arthritis and psychiatric disorders; results from a longitudinal population-based study. *J Psychosom Res* 2010 Feb;68(2):187-193. [doi: [10.1016/j.jpsychores.2009.05.011](https://doi.org/10.1016/j.jpsychores.2009.05.011)] [Medline: [20105702](https://pubmed.ncbi.nlm.nih.gov/20105702/)]
5. Axford J, Heron C, Ross F, Victor CR. Management of knee osteoarthritis in primary care: pain and depression are the major obstacles. *J Psychosom Res* 2008 May;64(5):461-467. [doi: [10.1016/j.jpsychores.2007.11.009](https://doi.org/10.1016/j.jpsychores.2007.11.009)] [Medline: [18440398](https://pubmed.ncbi.nlm.nih.gov/18440398/)]
6. Wang S, Ni G. Depression in osteoarthritis: current understanding. *Neuropsychiatr Dis Treat* 2022;18:375-389. [doi: [10.2147/NDT.S346183](https://doi.org/10.2147/NDT.S346183)] [Medline: [35237034](https://pubmed.ncbi.nlm.nih.gov/35237034/)]
7. Previtali D, Andriolo L, Di Laura Frattura G, Boffa A, Candrian C, Zaffagnini S, et al. Pain trajectories in knee osteoarthritis-a systematic review and best evidence synthesis on pain predictors. *J Clin Med* 2020 Sep 01;9(9):2828 [FREE Full text] [doi: [10.3390/jcm9092828](https://doi.org/10.3390/jcm9092828)] [Medline: [32882828](https://pubmed.ncbi.nlm.nih.gov/32882828/)]
8. Rathbun A, Shardell M, Stuart E, Yau M, Gallo J, Schuler M, et al. Pain severity as a mediator of the association between depressive symptoms and physical performance in knee osteoarthritis. *Osteoarthritis Cartilage* 2018 Nov;26(11):1453-1460 [FREE Full text] [doi: [10.1016/j.joca.2018.07.016](https://doi.org/10.1016/j.joca.2018.07.016)] [Medline: [30092262](https://pubmed.ncbi.nlm.nih.gov/30092262/)]
9. Rathbun AM, Stuart EA, Shardell M, Yau MS, Baumgarten M, Hochberg MC. Dynamic effects of depressive symptoms on osteoarthritis knee pain. *Arthritis Care Res* 2018 Jan 06;70(1):80-88 [FREE Full text] [doi: [10.1002/acr.23239](https://doi.org/10.1002/acr.23239)] [Medline: [28320048](https://pubmed.ncbi.nlm.nih.gov/28320048/)]
10. White DK, Neogi T, Nguyen UDT, Niu J, Zhang Y. Trajectories of functional decline in knee osteoarthritis: the Osteoarthritis Initiative. *Rheumatology* 2016 May;55(5):801-808 [FREE Full text] [doi: [10.1093/rheumatology/kev419](https://doi.org/10.1093/rheumatology/kev419)] [Medline: [26705330](https://pubmed.ncbi.nlm.nih.gov/26705330/)]
11. Kroenke K, Wu J, Bair MJ, Krebs EE, Damush TM, Tu W. Reciprocal relationship between pain and depression: a 12-month longitudinal analysis in primary care. *J Pain* 2011 Sep;12(9):964-973 [FREE Full text] [doi: [10.1016/j.jpain.2011.03.003](https://doi.org/10.1016/j.jpain.2011.03.003)] [Medline: [21680251](https://pubmed.ncbi.nlm.nih.gov/21680251/)]
12. Sharma A, Kudesia P, Shi Q, Gandhi R. Anxiety and depression in patients with osteoarthritis: impact and management challenges. *Open Access Rheumatol* 2016;8:103-113. [doi: [10.2147/OARRR.S93516](https://doi.org/10.2147/OARRR.S93516)] [Medline: [27843376](https://pubmed.ncbi.nlm.nih.gov/27843376/)]
13. Perruccio AV, Power JD, Evans HMK, Mahomed SR, Gandhi R, Mahomed NN, et al. Multiple joint involvement in total knee replacement for osteoarthritis: effects on patient-reported outcomes. *Arthritis Care Res* 2012 Jun;64(6):838-846. [doi: [10.1002/acr.21629](https://doi.org/10.1002/acr.21629)] [Medline: [22570306](https://pubmed.ncbi.nlm.nih.gov/22570306/)]
14. Rosemann T, Gensichen J, Sauer N, Laux G, Szecsenyi J. The impact of concomitant depression on quality of life and health service utilisation in patients with osteoarthritis. *Rheumatol Int* 2007 Jul 23;27(9):859-863. [doi: [10.1007/s00296-007-0309-6](https://doi.org/10.1007/s00296-007-0309-6)] [Medline: [17242902](https://pubmed.ncbi.nlm.nih.gov/17242902/)]

15. Gong L, Chen H. Descriptive analysis of the cost-effectiveness of depressed patients undergoing total knee arthroplasty: an economic decision analysis. *J Orthop Sci* 2014 Sep;19(5):820-826. [doi: [10.1007/s00776-014-0599-y](https://doi.org/10.1007/s00776-014-0599-y)] [Medline: [24996623](https://pubmed.ncbi.nlm.nih.gov/24996623/)]
16. Agarwal P, Sambamoorthi U. Healthcare expenditures associated with depression among individuals with osteoarthritis: post-regression linear decomposition approach. *J Gen Intern Med* 2015 Dec 20;30(12):1803-1811 [FREE Full text] [doi: [10.1007/s11606-015-3393-4](https://doi.org/10.1007/s11606-015-3393-4)] [Medline: [25990191](https://pubmed.ncbi.nlm.nih.gov/25990191/)]
17. Riddle DL, Kong X, Fitzgerald GK. Psychological health impact on 2-year changes in pain and function in persons with knee pain: data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage* 2011 Sep;19(9):1095-1101 [FREE Full text] [doi: [10.1016/j.joca.2011.06.003](https://doi.org/10.1016/j.joca.2011.06.003)] [Medline: [21723400](https://pubmed.ncbi.nlm.nih.gov/21723400/)]
18. Lin E, Tang L, Katon W, Hegel M, Sullivan M, Unützer J. Arthritis pain and disability: response to collaborative depression care. *Gen Hosp Psychiatry* 2006;28(6):482-486. [doi: [10.1016/j.genhosppsy.2006.08.006](https://doi.org/10.1016/j.genhosppsy.2006.08.006)] [Medline: [17088163](https://pubmed.ncbi.nlm.nih.gov/17088163/)]
19. Gleicher Y, Croxford R, Hochman J, Hawker G. A prospective study of mental health care for comorbid depressed mood in older adults with painful osteoarthritis. *BMC Psychiatry* 2011 Sep 12;11:147 [FREE Full text] [doi: [10.1186/1471-244X-11-147](https://doi.org/10.1186/1471-244X-11-147)] [Medline: [21910895](https://pubmed.ncbi.nlm.nih.gov/21910895/)]
20. Agarwal P, Pan X, Sambamoorthi U. Depression treatment patterns among individuals with osteoarthritis: a cross sectional study. *BMC Psychiatry* 2013 Apr 22;13:121 [FREE Full text] [doi: [10.1186/1471-244X-13-121](https://doi.org/10.1186/1471-244X-13-121)] [Medline: [23607696](https://pubmed.ncbi.nlm.nih.gov/23607696/)]
21. Cohen E, Lee YC. A mechanism-based approach to the management of osteoarthritis pain. *Curr Osteoporos Rep* 2015 Dec 30;13(6):399-406 [FREE Full text] [doi: [10.1007/s11914-015-0291-y](https://doi.org/10.1007/s11914-015-0291-y)] [Medline: [26419467](https://pubmed.ncbi.nlm.nih.gov/26419467/)]
22. Sayre E, Esdaile J, Kopec J, Singer J, Wong H, Thorne A, et al. Specific manifestations of knee osteoarthritis predict depression and anxiety years in the future: Vancouver Longitudinal Study of Early Knee Osteoarthritis. *BMC Musculoskelet Disord* 2020 Jul 16;21(1):467 [FREE Full text] [doi: [10.1186/s12891-020-03496-8](https://doi.org/10.1186/s12891-020-03496-8)] [Medline: [32677938](https://pubmed.ncbi.nlm.nih.gov/32677938/)]
23. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina* 2020 Sep 08;56(9):455 [FREE Full text] [doi: [10.3390/medicina56090455](https://doi.org/10.3390/medicina56090455)] [Medline: [32911665](https://pubmed.ncbi.nlm.nih.gov/32911665/)]
24. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* 2019 Mar 19;19(1):64 [FREE Full text] [doi: [10.1186/s12874-019-0681-4](https://doi.org/10.1186/s12874-019-0681-4)] [Medline: [30890124](https://pubmed.ncbi.nlm.nih.gov/30890124/)]
25. Nevitt MC, Felson DT, Lester G. The osteoarthritis initiative: protocol for the cohort study. National Institute of Mental Health Data Archive. 2006. URL: <https://nda.nih.gov/static/docs/StudyDesignProtocolAndAppendices.pdf> [accessed 2021-04-15]
26. Segal NA, Nevitt MC, Gross KD, Gross KD, Hietpas J, Glass NA, et al. The Multicenter Osteoarthritis Study: opportunities for rehabilitation research. *PM R* 2013 Aug;5(8):647-654 [FREE Full text] [doi: [10.1016/j.pmrj.2013.04.014](https://doi.org/10.1016/j.pmrj.2013.04.014)] [Medline: [23953013](https://pubmed.ncbi.nlm.nih.gov/23953013/)]
27. Smarr KL, Keefer AL. Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care Res* 2011 Nov;63(Suppl 11):S454-S466. [doi: [10.1002/acr.20556](https://doi.org/10.1002/acr.20556)] [Medline: [22588766](https://pubmed.ncbi.nlm.nih.gov/22588766/)]
28. WHO Consultation on Obesity. Obesity: preventing and managing the global epidemic: report of a WHO consultation. World Health Organization. 1999. URL: <https://apps.who.int/iris/handle/10665/42330> [accessed 2021-05-22]
29. Whelton P, Carey R, Aronow W, Casey D, Collins K, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension* 2018 Jun;71(6):1269-1324 [FREE Full text] [doi: [10.1161/HYP.000000000000066](https://doi.org/10.1161/HYP.000000000000066)] [Medline: [29133354](https://pubmed.ncbi.nlm.nih.gov/29133354/)]
30. Makizako H, Shimada H, Doi T, Tsutsumimoto K, Lee S, Hotta R, et al. Cognitive functioning and walking speed in older adults as predictors of limitations in self-reported instrumental activity of daily living: prospective findings from the Obu Study of Health Promotion for the Elderly. *Int J Environ Res Public Health* 2015 Mar 11;12(3):3002-3013 [FREE Full text] [doi: [10.3390/ijerph120303002](https://doi.org/10.3390/ijerph120303002)] [Medline: [25768239](https://pubmed.ncbi.nlm.nih.gov/25768239/)]
31. Xu Y, Hong K, Tsujii J, Chang EI. Feature engineering combined with machine learning and rule-based methods for structured information extraction from narrative clinical discharge summaries. *J Am Med Inform Assoc* 2012 Sep 01;19(5):824-832 [FREE Full text] [doi: [10.1136/amiajnl-2011-000776](https://doi.org/10.1136/amiajnl-2011-000776)] [Medline: [22586067](https://pubmed.ncbi.nlm.nih.gov/22586067/)]
32. Weng W. Machine learning for clinical predictive analytics. In: Celi L, Majumder M, Ordóñez P, Osorio J, Paik K, Somai M, editors. *Leveraging data science for global health*. Cham: Springer; 2020. URL: [https://link.springer.com/chapter/10.1007/978-3-030-47994-7\\_12](https://link.springer.com/chapter/10.1007/978-3-030-47994-7_12)
33. Chen JJ, Tsai CA, Moon H, Ahn H, Young JJ. The use of decision threshold adjustment in classification for cancer prediction. Penn State University. 2005. URL: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.5813&rep=rep1&type=pdf> [accessed 2021-04-22]
34. RStudio: Integrated Development for R. 2020. URL: <https://www.rstudio.com/> [accessed 2021-05-28]
35. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*, 3rd edition. Hoboken, NJ: Wiley; 2013.

36. The R stats package (version 3.6.2). RDocumentation. URL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2> [accessed 2021-05-28]
37. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc B* 2018 Dec 05;58(1):267-288. [doi: [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x)]
38. Lasso and elastic-net regularized generalized linear models (glmnet package). RDocumentation. URL: <https://www.rdocumentation.org/packages/glmnet/versions/4.1-1> [accessed 2021-05-28]
39. Quinlan JR. Induction of decision trees. *Mach Learn* 1986 Mar;1(1):81-106. [doi: [10.1007/bf00116251](https://doi.org/10.1007/bf00116251)]
40. rpart package. RDocumentation. URL: <https://www.rdocumentation.org/packages/rpart/versions/4.1-15> [accessed 2021-05-28]
41. Breiman L. Random forests. *Mach Learn* 2001;45:5-32. [doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)]
42. randomForest package. RDocumentation. URL: <https://www.rdocumentation.org/packages/randomForest/versions/4.6-14> [accessed 2021-05-28]
43. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurobot* 2013;7:21. [doi: [10.3389/fnbot.2013.00021](https://doi.org/10.3389/fnbot.2013.00021)] [Medline: [24409142](https://pubmed.ncbi.nlm.nih.gov/24409142/)]
44. gbm: Generalized Boosted Regression Modeling (GBM). RDocumentation. URL: <https://www.rdocumentation.org/packages/gbm/versions/2.1.8/topics/gbm> [accessed 2021-05-28]
45. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 2003 Jul;29(7):1043-1051. [doi: [10.1007/s00134-003-1761-8](https://doi.org/10.1007/s00134-003-1761-8)] [Medline: [12734652](https://pubmed.ncbi.nlm.nih.gov/12734652/)]
46. Anderson M, Anderson S. How should AI be developed, validated, and implemented in patient care? *AMA J Ethics* 2019 Feb 01;21(2):E125-E130 [FREE Full text] [doi: [10.1001/amajethics.2019.125](https://doi.org/10.1001/amajethics.2019.125)] [Medline: [30794121](https://pubmed.ncbi.nlm.nih.gov/30794121/)]
47. Japkowicz N, Stephen S. The class imbalance problem: a systematic study. *Intell Data Anal* 2002 Nov 15;6(5):429-449. [doi: [10.3233/ida-2002-6504](https://doi.org/10.3233/ida-2002-6504)]
48. Diprose W, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. *J Am Med Inform Assoc* 2020 Apr 01;27(4):592-600 [FREE Full text] [doi: [10.1093/jamia/ocz229](https://doi.org/10.1093/jamia/ocz229)] [Medline: [32106285](https://pubmed.ncbi.nlm.nih.gov/32106285/)]
49. Zheng S, Tu L, Cicuttini F, Zhu Z, Han W, Antony B, et al. Depression in patients with knee osteoarthritis: risk factors and associations with joint symptoms. *BMC Musculoskelet Disord* 2021 Jan 07;22(1):40 [FREE Full text] [doi: [10.1186/s12891-020-03875-1](https://doi.org/10.1186/s12891-020-03875-1)] [Medline: [33413273](https://pubmed.ncbi.nlm.nih.gov/33413273/)]
50. Kim KW, Han JW, Cho HJ, Chang CB, Park JH, Lee JJ, et al. Association between comorbid depression and osteoarthritis symptom severity in patients with knee osteoarthritis. *J Bone Joint Surg Am* 2011 Mar 16;93(6):556-563. [doi: [10.2106/JBJS.I.01344](https://doi.org/10.2106/JBJS.I.01344)] [Medline: [21411706](https://pubmed.ncbi.nlm.nih.gov/21411706/)]
51. Georgiev T, Angelov AK. Modifiable risk factors in knee osteoarthritis: treatment implications. *Rheumatol Int* 2019 Jul;39(7):1145-1157. [doi: [10.1007/s00296-019-04290-z](https://doi.org/10.1007/s00296-019-04290-z)] [Medline: [30911813](https://pubmed.ncbi.nlm.nih.gov/30911813/)]
52. Monroe SM, Slavich GM, Gotlib IH. Life stress and family history for depression: the moderating role of past depressive episodes. *J Psychiatr Res* 2014 Feb;49:90-95 [FREE Full text] [doi: [10.1016/j.jpsychires.2013.11.005](https://doi.org/10.1016/j.jpsychires.2013.11.005)] [Medline: [24308926](https://pubmed.ncbi.nlm.nih.gov/24308926/)]
53. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014 Mar 19;14(1):40 [FREE Full text] [doi: [10.1186/1471-2288-14-40](https://doi.org/10.1186/1471-2288-14-40)] [Medline: [24645774](https://pubmed.ncbi.nlm.nih.gov/24645774/)]
54. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006 Feb 07;144(3):201-209. [doi: [10.7326/0003-4819-144-3-200602070-00009](https://doi.org/10.7326/0003-4819-144-3-200602070-00009)] [Medline: [16461965](https://pubmed.ncbi.nlm.nih.gov/16461965/)]
55. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988 Dec;15(12):1833-1840. [Medline: [3068365](https://pubmed.ncbi.nlm.nih.gov/3068365/)]
56. 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](https://doi.org/10.1097/00005650-199603000-00003)] [Medline: [8628042](https://pubmed.ncbi.nlm.nih.gov/8628042/)]
57. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1(3):385-401. [doi: [10.1177/014662167700100306](https://doi.org/10.1177/014662167700100306)]
58. Shin D, Lee KJ, Adeluwa T, Hur J. Machine learning-based predictive modeling of postpartum depression. *J Clin Med* 2020 Sep 08;9(9):2899 [FREE Full text] [doi: [10.3390/jcm9092899](https://doi.org/10.3390/jcm9092899)] [Medline: [32911726](https://pubmed.ncbi.nlm.nih.gov/32911726/)]
59. Wang S, Pathak J, Zhang Y. Using electronic health records and machine learning to predict postpartum depression. *Stud Health Technol Inform* 2019 Aug 21;264:888-892. [doi: [10.3233/SHTI190351](https://doi.org/10.3233/SHTI190351)] [Medline: [31438052](https://pubmed.ncbi.nlm.nih.gov/31438052/)]
60. Cvetković J. Breast cancer patients' depression prediction by machine learning approach. *Cancer Invest* 2017 Sep 14;35(8):569-572. [doi: [10.1080/07357907.2017.1363892](https://doi.org/10.1080/07357907.2017.1363892)] [Medline: [28872366](https://pubmed.ncbi.nlm.nih.gov/28872366/)]
61. Nemesure MD, Heinz MV, Huang R, Jacobson NC. Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. *Sci Rep* 2021 Jan 21;11(1):1980. [doi: [10.1038/s41598-021-81368-4](https://doi.org/10.1038/s41598-021-81368-4)] [Medline: [33479383](https://pubmed.ncbi.nlm.nih.gov/33479383/)]
62. Choi J, Choi J, Choi WJ. Predicting depression among community residing older adults: a use of machine learning approach. *Stud Health Technol Inform* 2018;250:265. [Medline: [29857458](https://pubmed.ncbi.nlm.nih.gov/29857458/)]
63. Stubbs B, Aluko Y, Myint PK, Smith TO. Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis. *Age Ageing* 2016 Mar;45(2):228-235. [doi: [10.1093/ageing/afw001](https://doi.org/10.1093/ageing/afw001)] [Medline: [26795974](https://pubmed.ncbi.nlm.nih.gov/26795974/)]

## Abbreviations

**AUC:** area under the receiver operating characteristic curve  
**CES-D:** Center for Epidemiological Studies Depression Scale  
**GBM:** gradient boosting machine  
**LASSO:** least absolute shrinkage and selection operator  
**ML:** machine learning  
**MOST:** Multicenter Osteoarthritis Study  
**OA:** osteoarthritis  
**OAI:** Osteoarthritis Initiative  
**PASE:** Physical Activity Scale for the Elderly  
**SF-12:** 12-item Short Form Health Survey  
**WOMAC:** Western Ontario and McMaster Universities Osteoarthritis Index

*Edited by A Mavragani; submitted 03.01.22; peer-reviewed by I Kim, M Pritchard; comments to author 26.05.22; revised version received 31.07.22; accepted 09.08.22; published 13.09.22.*

*Please cite as:*

Nowinka Z, Alagha MA, Mahmoud K, Jones GG

*Predicting Depression in Patients With Knee Osteoarthritis Using Machine Learning: Model Development and Validation Study*

*JMIR Form Res 2022;6(9):e36130*

URL: <https://formative.jmir.org/2022/9/e36130>

doi: [10.2196/36130](https://doi.org/10.2196/36130)

PMID: [36099008](https://pubmed.ncbi.nlm.nih.gov/36099008/)

©Zuzanna Nowinka, M Abdulhadi Alagha, Khadija Mahmoud, Gareth G Jones. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 13.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Examining the Delivery of a Tailored Chinese Mind-Body Exercise to Low-Income Community-Dwelling Older Latino Individuals for Healthy Aging: Feasibility and Acceptability Study

Yan Du<sup>1</sup>, RN, MPH, PhD; Neela Patel<sup>2</sup>, MPH, MD; Arthur Hernandez<sup>3</sup>, PhD; Maria Zamudio-Samano<sup>4</sup>, AAS; Shiyu Li<sup>1</sup>, MS; Tianou Zhang<sup>5</sup>, PhD; Roman Fernandez<sup>6</sup>, MS; Byeong Yeob Choi<sup>6</sup>, PhD; William M Land<sup>5</sup>, PhD; Sarah Ullevig<sup>7</sup>, PhD; Vanessa Estrada Coats<sup>8</sup>, LD, RDN, MS; Jessh Mondesir Mavoungou Moussavou<sup>8</sup>, BS; Deborah Parra-Medina<sup>9</sup>, MPH, PhD; Zenong Yin<sup>8</sup>, PhD

<sup>1</sup>Center on Smart and Connected Health Technology, School of Nursing, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

<sup>2</sup>Department of Family and Community Medicine, School of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

<sup>3</sup>Dreeben School of Education, University of the Incarnate Word, San Antonio, TX, United States

<sup>4</sup>Alamo Colleges District-North Vista College, San Antonio, TX, United States

<sup>5</sup>Department of Kinesiology, University of Texas at San Antonio, San Antonio, TX, United States

<sup>6</sup>Department of Population Health Science, School of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

<sup>7</sup>College for Health, Community and Policy, University of Texas at San Antonio, San Antonio, TX, United States

<sup>8</sup>Department of Public Health, University of Texas at San Antonio, San Antonio, TX, United States

<sup>9</sup>Latino Research Institute, Latino Studies, University of Texas at Austin, Austin, TX, United States

**Corresponding Author:**

Yan Du, RN, MPH, PhD

Center on Smart and Connected Health Technology

School of Nursing

University of Texas Health Science Center at San Antonio

7703 Floyd Curl Dr

San Antonio, TX, 78229

United States

Phone: 1 5044931202

Email: [duy@uthscsa.edu](mailto:duy@uthscsa.edu)

## Abstract

**Background:** Older Latino individuals are disproportionately affected by various chronic conditions including impairments in physical and cognitive functions, which are essential for healthy aging and independent living.

**Objective:** This study aimed to evaluate the feasibility and acceptability of FITxOlder, a 12-week mind-body exercise program, in community-dwelling low-income, predominantly older Latino individuals, and assess its preliminary effects on health parameters relevant to healthy aging and independent living.

**Methods:** This 12-week, single-arm, stage 1B feasibility study had a pre- and poststudy design. A total of 13 older adults (mean age 76.4, SD 7.9 years; 11/13, 85% Latino) of a congregate meal program in a senior center were enrolled. FITxOlder was a tailored Chinese mind-body exercise program using Five Animal Frolics led by a bilingual community health worker (CHW) participating twice a week at the senior center and facilitated by mobile health technology for practice at home, with incrementally increasing goals moving from once a week to at least 3 times a week. The feasibility and acceptability of the study were examined using both quantitative and qualitative data. Healthy aging-related outcomes (eg, physical and cognitive function) were assessed using paired 2-tailed *t* tests. Qualitative interview data were analyzed using thematic analysis.

**Results:** The attendance rate for the 24 exercise sessions was high (22.7/24, 95%), ranging from 93% (1.8/2) to 97% (1.9/2) over the 12 weeks. Participants were compliant with the incremental weekly exercise goals, with 69.2% (9/13) and 75.0% (9/12) meeting the home and program goals in the last 4 weeks, respectively. Approximately 83% (10/12) to 92% (11/12) of the



participants provided favorable feedback on survey questions regarding the study and program implementation, such as program content and support, delivery by the CHW, enjoyment and appeal of the Five Animal Frolics, study burden and incentives, and safety concerns. The qualitative interview data revealed that FITxOlder was well accepted; participants reported enjoyment and health benefits and the desire to continue to practice and share it with others. The 5-time sit-to-stand test (mean change at posttest assessment = -1.62;  $P < .001$ ; Cohen  $d = 0.97$ ) and 12-Item Short Form Health Survey physical component scores (mean change at post intervention = 5.71;  $P = .01$ ; Cohen  $d = 0.88$ ) exhibited changes with large effect sizes from baseline to 12 weeks; the other parameters showed small or medium effect sizes.

**Conclusions:** The research findings indicated that the CHW-led and mobile health-facilitated Chinese qigong exercise program is feasible and acceptable among low-income Latino older adults. The trending health benefits of the 12-week FITxOlder program suggest it is promising to promote physical activity engagement in underserved older populations to improve health outcomes for healthy aging and independent living. Future research with larger samples and longer interventions is warranted to assess the health benefits and suitability of FITxOlder.

(*JMIR Form Res* 2022;6(9):e40046) doi:[10.2196/40046](https://doi.org/10.2196/40046)

## KEYWORDS

qigong; mind-body exercise; Five Animal Frolics; health technology; older adults; Hispanics; Latinos; low-income; healthy aging; aging in place; independent living

## Introduction

### Background

Healthy aging is “the process of developing and maintaining the functional ability that enables well-being in older age” [1]. Being able to live independently is one of the key characteristics of healthy aging [2]. However, aging is accompanied by a variety of chronic conditions, and healthy aging is challenged by physical and cognitive function decline and, consequently, a lack of independence and compromised quality of life [3]. Approximately 75% of the diversity in the capacity and circumstances of healthy aging in older age results from the cumulative impact of advantages and disadvantages (eg, income and education) across people’s lives, and race and ethnicity are among the many contributing factors [1].

As the largest and most rapidly growing racial and ethnic minority in the United States [3,4], individuals aged  $\geq 65$  years of Latino or Hispanic origin have increasingly experienced health disparities in the aging process despite the noted Latino health paradox [5,6]. Older Latino individuals are more likely to have deteriorating physical and cognitive functions for independent living because of disabilities, poor mental health, and burdens of chronic disease from a lifetime exposure to poor conditions of daily life, poverty, discrimination, and substandard health care compared with their White counterparts [4,5,7,8]. This health disparity has contributed to longer durations of disabilities, a lower quality of life, and higher costs of health care in older Latino individuals while placing an increased burden on family members who provide their care during late life [4,5,7,8]. Despite urgent calls to develop culturally tailored and sustainable aging programs, research remains scarce on translating evidence-based practices to promote healthy aging in older Latino individuals who live in underserved communities [9-11].

As a distinctive form of complementary and alternative medicine [12], mind-body exercise refers to meditative movements that combine body movement, controlled breathing, and mental focus (or spirituality) to improve the attributes of physical fitness

(strength, endurance, balance, and flexibility) and health-related outcomes [13,14]. Qigong, tai chi, and yoga are a few common forms of mind-body exercises that are rooted in Asian culture that have been accepted and practiced in Western countries in recent decades [15-17]. A well-established body of research using randomized controlled trials has demonstrated the clinically significant benefits of qigong, tai chi, and yoga on a variety of health outcomes, including but not limited to physical and cognitive function, quality of life, mental disorders, and chronic pain in diverse clinical and nonclinical populations in Western and non-Western countries [18-22]. Unlike traditional Western exercise, qigong exercise is characterized as low- to moderate-intensity, community-oriented with no or low cost, low-demand regarding space and equipment, and safe for all age groups and health conditions in Asian countries [22,23]. Recently, physicians in the United States and other Western countries have increased the use of the mind-body exercise for managing chronic health conditions and mental disorders following a holistic or integrative health framework. However, the uptake of mind-body exercises in Western countries is mostly limited to adults of younger age and higher socioeconomic status owing to limited availability and lack of awareness of health benefits compared with other forms of alternative health practices in the United States [16,17,24,25].

Although qigong and tai chi are often used interchangeably, qigong was generally used for the purpose of illness prevention and treatment by focusing on combining the physical and spiritual nature of the routines without the push for physical exhaustion, whereas tai chi had an origin in martial arts [26]. *Five Animal Frolics* or Five Animal Play (*Wu Qin Xi* in Chinese) is a traditional Chinese qigong that mimics the movements and spirit of a tiger, deer, bear, monkey, and crane to cultivate the energy emitted from the 5 animals and create harmony within all parts of the body and with the universe [26,27]. The playful and easy-to-learn movements of Five Animal Frolics can be practiced at different levels of physical exertion and range of movements to control the difficulty and stress loads of the exercise and, therefore, appeal to diverse groups, especially older adults [28].

Although evidence shows that qigong exercises can improve physical and cognitive function and quality of life and reduce fall risk, depression, and anxiety in older adults [21,22,29-32], methodological issues limit the scientific rigor and generalizability of the findings in older underserved racial and ethnic minority adults in the United States [22]. For example, most studies were conducted in China or with participants of higher socioeconomic status in Western countries. Another concern is the failure to address the spiritual, social, and cultural aspects of qigong exercise that can influence the uptake and maintenance of the mind-body exercises in older adults [28,33,34]. Although the concept of qi or energy is central in the teaching and practice of qigong, its meaning remains controversial among Western practitioners and researchers, and its acceptance has not been evaluated in studies conducted in Western populations [35]. This failure has contributed to the hesitation in recommending qigong and other mind-body exercises as part of holistic health care [36]. Finally, few studies have examined the barriers that limit the access to and availability and sustainability of mind-body exercise, such as transportation, family-friendliness, language, technology, and program cost in older underserved minority adults in the United States [33].

## Objectives

This paper reports the findings of a stage 1B feasibility pilot study [37] of a refined 12-week community-based and tailored Chinese qigong program using the Five Animal Frolics, “Function Improvement Exercises for Older Sedentary Community-Dwelling Latino Residents (FITxOlder),” to promote healthy aging in community-dwelling Latino older adults [38]. The primary purposes of the study were to (1) evaluate the feasibility of the intervention protocol in terms of program participation, fidelity of implementation, feasibility of the data collection protocol, progression of the stepped exercise program, and study-related adverse events and (2) examine the participants’ satisfaction with and feedback on the delivery of the 12-week program, cultural and age appropriateness of the intervention content and delivery, recruitment and retention practices, data collection protocol, and use of technologies for facilitating and supporting program delivery. The study also explored the changes in healthy aging–related outcomes in response to the exercise intervention.

## Methods

### Study Design and Sample

We conducted a single-arm stage 1B feasibility study with a pre- and posttest design to evaluate the feasibility and acceptability of FITxOlder, a 12-week healthy aging program, in older adults who were enrollees of a congregate meal program in a senior center in San Antonio, Texas. Most participants in the selected center were older Latino individuals. All the program participants were eligible to take part in the study if they were able to exercise in a standing position (with or without an assistive device), owned a cell phone, and agreed not to participate in other exercise programs. Those who planned to leave San Antonio during the 3-month study period were excluded. The senior center staff distributed the recruitment

flyers to potential program participants. Those who were interested and met the study eligibility criteria attended an orientation session to learn the details of the study.

### Ethical Considerations

All participants completed a consent form to take part in the study. Recruitment information and consent forms were available in English and Spanish. The participants received up to US \$90 (US \$20 at baseline, US \$30 at midtest assessment, and US \$40 at posttest assessment) in grocery gift cards to compensate for their time for participation in data collection. The study protocol was approved by the University of Texas at San Antonio Institutional Review Board (protocol FY20-21-259).

### Intervention Program Description

FITxOlder was a community-based healthy aging program that was delivered by a trained community health worker (CHW) in a senior center with support and coordination from the senior center director. The FITxOlder intervention in this study was based on the protocol for a previous 3-arm feasibility study that explored the acceptance of qigong exercises and developed and piloted a healthy aging program using Five Animal Frolics among low-income Latino older adults. The original feasibility study included 2 phases: (1) planning and developing the intervention protocol based on input from the target population and (2) pilot-testing the feasibility and acceptability of the developed intervention protocol. The movement routines of Five Animal Frolics used in our study consisted of an opening routine, tiger routine 1 (raising tiger paws), tiger routine 2 (seizing the prey), deer routine 1 (colliding with the antlers), deer routine 2 (running like a deer), monkey routine 1 (lifting the monkey’s paws), monkey routine 2 (picking fruits), bear routine 1 (rotating the waist like a bear), bear routine 2 (swaying like a bear), crane routine 1 (stretching upward), crane routine 2 (flying like a crane), and closing routine. The routines are usually performed following a 13-minute narrated audio or video with background music embedded with animal sounds [27]. In the original pilot, one group learned the official version of Five Animal Frolics, the second group learned a modified version of Five Animal Frolics, and the third group received active control. The 4 components of the original FITxOlder were a CHW-led group session (didactic education, instruction based on the learning needs of older learners, and supervised practice), goal-based home practice (weekly exercise goals), a 3-stepped intervention with different instruction foci, and technology-facilitated support (tablet for showing the exercise video, SMS text message reminders, and follow-up phone calls), as described in Table 1. The weekly exercise goals increased from a minimum of 3 times a week to a minimum of 5 times a week by reaching an equivalent of  $\geq 150$  minutes of light to moderate physical activity (PA) per week. We also produced a 32-minute video (5-minute warm-up and replay of 13-minute movement routines of the official and modified Five Animal Frolics with a 1-minute break) to support and facilitate learning and practice at home. Unfortunately, the COVID-19 pandemic disrupted the implementation of the study protocol and prevented a full evaluation of the feasibility and acceptability of the pilot intervention. The rationale, development, and results

of the original FITxOlder pilot program have been published elsewhere [38].

The protocol for the 12-week FITxOlder intervention in this study was refined using the findings from the original pilot intervention and poststudy participant focus group discussion to increase the program's feasibility and acceptability. First, we aligned the conceptual framework of the intervention with the information-motivation-behavioral skills model that informed factors influencing the uptake and maintenance of a health practice (ie, participants' learning and practice of Five Animal Frolics) [39]. FITxOlder provided the participants with knowledge of the background, healthy aging benefits, inner workings, and safety of qigong and Five Animal Frolics to increase their understanding and acceptance of the program. Motivation was facilitated by using a group format for social

support, modeling by peers, and SMS text messaging and monitoring by a CHW and center staff. FITxOlder also focused on developing behavioral skills (goal setting and self-efficacy for practicing Five Animal Play) to engage the informed and motivated participants in the program and achieve the weekly exercise goals.

Second, the participants learned the modified Five Animal Frolics movements following the stepped instruction progression used in the original pilot (Table 1). In week 10, we introduced an advanced version of the modified Five Animal Frolics that increased the level of difficulty of the movements and physical exertion by increasing the range of movements and demand for balance (eg, on 1 foot). A community videographer made a new video of the modified Five Animal Frolics and a video of the advanced version of the routines for the study participants.

**Table 1.** Components of the FITxOlder program<sup>a</sup>.

Weeks	Intervention activities
1 to 4 (step 1)	<ul style="list-style-type: none"> <li>• Biweekly group sessions: attending two 60-minute group sessions led by a CHW<sup>b</sup> each week</li> <li>• Weekly home exercise goals: practice of Five Animal Frolics at least once at home following a video on a tablet</li> <li>• Weekly program exercise goals: practice of Five Animal Frolics at least three times (twice at the center and at least once at home)</li> <li>• Didactic education: introduction of Five Animal Frolics and qigong to the participants</li> <li>• Instruction focus: teaching abdominal breathing; teaching choreography of the modified movement routines</li> <li>• Participant support: weekly SMS text message reminder to perform the exercise at home and follow-up contact call by the CHW or senior center after missing a group session</li> </ul>
5 to 8 (step 2)	<ul style="list-style-type: none"> <li>• Biweekly group sessions: attending two 60-minute group sessions led by a CHW each week</li> <li>• Weekly home exercise goals: practice of Five Animal Frolics at least twice at home following a video on a tablet</li> <li>• Weekly program exercise goals: practice of Five Animal Frolics at least four times (twice at the center and at least twice at home)</li> <li>• Didactic education: visualizing the animals while performing the routine, mimicking the animals' movements and spirituality, and discussing the correspondence between the movement routines and the health-related elements</li> <li>• Instruction focus: teaching blending movements and breathing</li> <li>• Participant support: weekly SMS text message reminder to perform the exercise at home and follow-up contact call by the CHW or senior center after missing a group session</li> </ul>
9 to 12 (step 3)	<ul style="list-style-type: none"> <li>• Biweekly group sessions: attending two 60-minute group sessions led by a CHW each week</li> <li>• Weekly home exercise goals: practice of Five Animal Frolics at least three times at home following a video on a tablet</li> <li>• Weekly program exercise goals: practice of Five Animal Frolics at least five times (twice at the center and at least three times at home)</li> <li>• Didactic education: teaching the consciousness of the present moment and energy, discussing the connections between the movements and the animals' spirituality, and stressing the importance of combining the movements and breathing</li> <li>• Instruction focus: teaching the blending of movements, breathing, and the animals' spirit into "one" and teaching the advanced version of Five Animal Frolics</li> <li>• Participant support: weekly SMS text message reminder to perform the exercise at home and follow-up contact call by the CHW or senior center after missing a group session</li> </ul>

<sup>a</sup>The intervention protocol developed for the original pilot study.

<sup>b</sup>CHW: community health worker.

Third, in response to the feedback regarding the participants' interest in learning the background of qigong, we decided to introduce basic information on qigong and Five Animal Frolics: the meaning of qi, history of Five Animal Frolics, symbolism (spiritual or cultural meanings) of the 5 animals, health-related elements associated with movement routines, and consciousness of the present moment. Exposing the participants to the basic background of qigong also allowed us to explore the appropriateness and acceptance of the cultural and spiritual aspects as part of qigong or Asian mind-body teaching.

Finally, we refined the group sessions to offer a supportive learning environment to promote engagement and a sense of mastery in the study participants who were older adult learners

following the principles of geragogy [40-42]. The instruction strategies used by the CHW included (1) teaching the movement routine with a whole-part-whole learning model; (2) offering opportunities for active hands-on learning with repeated exposure to the content; (3) developing rapport with the participants, communicating with respect, and recognizing individual differences; (4) offering the background and reasons for learning a task at the early stage of learning; (5) promoting and supporting self-directed and self-paced learning with respect to the participants' life experience; and (6) providing regular feedback and focusing on small progress with positive reinforcement and support. The 60-minute session was divided into check-in and support, warm-up, supervised practice,

instruction, and administrative time that allowed adequate time for participant engagement and support, as well as a minimum of 30 minutes dedicated to PA (Multimedia Appendix 1).

### Study Measures and Data Collection Procedures

There were 3 types of measures, which offered data to evaluate the feasibility and acceptability of the study, intervention protocols, and changes in healthy aging-related outcomes in

response to the exercise intervention. Table 2 describes the measures and criteria used to evaluate the study feasibility. In addition, the CHW completed a group exercise session evaluation form to document reasons for missing the session, check the completion of planned class activities and participation, and report problems with lesson delivery at the end of each group session.

**Table 2.** Measures of feasibility.

Measure	Measurement procedure	Success criteria
Participant retention: completion rate of midtest and posttest assessment	Percentage of participants who completed the midtest and posttest assessment	≥80%
Participation: days attending the biweekly group sessions	Percentage of participants who attended the biweekly group sessions at the senior center (2 days per week) based on the attendance records kept by the CHW <sup>a</sup>	≥70%
Participation: days of practicing Five Animal Frolics using the 32-minute video at home following the weekly home exercise goals	Percentage of participants who achieved the weekly home exercise goals (≥1 day per week in weeks 1-4, ≥2 days per week in weeks 5-8, and ≥3 days per week in weeks 9-12) based on the reports from the weekly exercise logs kept by the participants	≥50%
Participation: total number of days of practicing the Five Animal Frolics exercise at the center and at home following the weekly program exercise goals	Percentage of participants who achieved the weekly exercise goals (≥3 days per week in weeks 1-4, ≥4 days per week in weeks 5-8, and ≥5 days per week in weeks 9-12) based on the attendance records and the weekly exercise logs	≥50%
Feasibility of the data collection protocol	Percentage of participants who completed all aging-related outcome measures at the baseline, 6-week, and 12-week assessments	80%
Study-related adverse events	Number of unanticipated adverse events that were related to the study in weeks 1 to 12 based on reports from the CHW and participants	None

<sup>a</sup>CHW: community health worker.

After completing the exercise in weeks 1 to 12, participants completed the Exercise-Induced Feeling Inventory to report feelings regarding revitalization, tranquility, positive engagement, and physical exhaustion as reactivity to the exercise on a 5-point scale from 0 to 4, where 0 stood for “do not feel at all” and 4 stood for “feel very strongly.” To document whether the meditative state of mind-body connection improved over time, participants completed a modified version of the Meditative Movement Inventory (MMI) [43] to report their perceived state in 2 dimensions (breath focus and meditative connection) on a 5-point scale (5 for *all the time*, 4 for *very frequently*, 3 for *occasionally rarely*, 2 for *very rarely*, and 1 for *never*) at weeks 2, 5, 7, 9, and 12. Higher scores indicated a higher level of meditative state. Finally, we created an observation form to evaluate the level of proficiency of the participants performing each routine of Five Animal Frolics in 3 dimensions (smooth and fluid movement, consistency of movement, and accurate imitation of movement) on a 5-point scale where 1 stood for *not at all*, 2 stood for *poor*, 3 stood for *fair*, 4 stood for *good*, and 5 stood for *excellent*. The evaluation was conducted by 2 observers monitoring 1 participant per routine who was selected randomly. The observations were repeated during the first and second rounds of the Five Animal Frolics exercise in 1 group session in weeks 5, 7, 9, and 12. The average of the scores from the 2 rounds was used. Higher scores indicated a higher level of proficiency.

To assess the acceptability of the intervention, the participants completed a poststudy survey to evaluate their satisfaction with the program content and delivery and provide feedback on the

study protocol related to study burdens and incentives and exercise safety. Facilitation using digital technologies (using videos for modeling, tablets for showing the videos, and SMS text messages) was also evaluated. We expected that at least two-thirds of the participants would have a favorable response to the questions in the evaluation survey. We also conducted a focus group discussion with 7 study participants led by an English-Spanish bilingual facilitator to gather feedback on the delivery and content of the FITxOlder program 3 weeks after the completion of the study. The 7 participants were selected conveniently as whoever was available on a selected date when most participants could attend. The focus group discussion was guided by a predesigned interview guide, including topics such as *what challenges did you encounter when you practiced the exercise?* Probing questions were asked wherever appropriate. The discussion lasted approximately 60 minutes. The session was audio recorded and transcribed for analysis.

As part of the feasibility study, we collected data to explore the changes in healthy aging-related physical, cognitive, physiological, psychosocial, and behavioral outcomes in response to the exercise intervention (Table 3) at midtest assessment in week 6 and posttest assessment in week 13. The participants received written instructions 3 days in advance to prepare for the measurements. Trained bilingual research assistants and research faculty investigators conducted the physical (40 minutes), cognitive (30 minutes), and psychosocial (30 minutes) measurement of the participants in one 2-hour session at the senior center during the morning hours. All psychosocial measures were available in English and Spanish.

At the end of the measurement session, the participants received a gift card for taking part in the data collection. The participants also received a report of their weight, blood pressure, and

physical function tests over the 3 measurement time points at the end of the study.

**Table 3.** Description of outcome measures.

Outcome (healthy aging–related outcomes) and measure	Measurement procedure
Anthropometric measures and blood pressure	<ul style="list-style-type: none"> <li>Height (cm) and weight (kg) were measured 2 times without shoes. The average was used.</li> <li>Systolic and diastolic blood pressure were measured 2 times after a 5-minute rest. The average was used.</li> </ul>
Physical function	<ul style="list-style-type: none"> <li>Physical function was assessed with a battery of physical function tests consisting of the 5XSTS<sup>a</sup>, 50-foot FWT<sup>b</sup>, hand grip test, 6MWT<sup>c</sup>, and FLRT<sup>d</sup> [44].</li> </ul>
Cognitive function	<ul style="list-style-type: none"> <li>The Clock Drawing Test is a measure of visual-spatial abilities and cognitive function and has been used as a screen for cognitive impairment and dementia [45]. Participants are asked to draw the face of a clock with the hands set to “ten minutes past 11 o’clock.” Higher scores indicate high cognitive function.</li> <li>Performance on the Trail Making Test Part A and B is linked to visual search speed, speed of cognitive processing, mental flexibility, and executive functioning [46]. Participants are asked to draw a line to connect circles numbered 1 to 25 in ascending order in Part A and to connect the circles in an ascending order while alternating between numbers and letters (ie, 1-A-2-B-3-C) in Part B as quickly as possible. Longer time (seconds) to complete the task indicates lower level of cognitive function.</li> </ul>
Psychosocial responses	<ul style="list-style-type: none"> <li>Participants completed the SF-12<sup>e</sup> [47] to generate a physical component score and mental component score of quality of life.</li> <li>Participants completed the BPI<sup>f</sup> to report the level of perceived pain in 2 domains: pain severity (4 items) and pain interference with life (7 items) [48].</li> <li>The participants completed the Sleep Disturbance Short Form 8a (8 items) [49] to report the quality of their sleep.</li> </ul>
Behavioral response	<ul style="list-style-type: none"> <li>Participants completed the basic ADL<sup>g</sup> subscale (5 items) and intermediate ADL subscale (4 items) of the Functional Status Questionnaire [50].</li> <li>The participants wore a GENEActiv triaxial accelerometer (Activinsights) on the nondominant wrist for 7 days. Time (hours per day) spent in sleep, sedentary activities, and light physical activity was estimated using the R package GGIR (version 1.10-7; R Foundation for Statistical Computing) [51] in the R environment (version 3.6.1).</li> </ul>
Biomarker responses (data not reported)	<ul style="list-style-type: none"> <li>A saliva sample was collected after a 12-hour fast. The participants also provided samples of saliva collected at home using a saliva collection kit.</li> </ul>

<sup>a</sup>5XSTS: 5-time sit-to-stand test.

<sup>b</sup>FWT: fast walk test.

<sup>c</sup>6MWT: 6-minute walk test.

<sup>d</sup>FLRT: forward lean reach test.

<sup>e</sup>SF-12: 12-Item Short Form Health Survey.

<sup>f</sup>BPI: Brief Pain Inventory.

<sup>g</sup>ADL: activities of daily living.

## Data Analyses

The characteristics of the study participants were analyzed using descriptive analyses and are presented as means and SDs for continuous variables and percentages for categorical variables. We adopted descriptive statistics to assess the measures of feasibility and acceptability. Participants’ responses in the focus group discussion were processed using qualitative content analysis [52] to identify themes related to experiences of participating in the FITxOlder program. We also conducted a paired *t* test (2-tailed) to assess the responsiveness (ie, changes) in the healthy aging–related outcome variables from baseline to midtest and posttest assessments. The Cohen effect size coefficient (Cohen *d*) was used to evaluate the changes in each outcome (Cohen *d*=0.2 for small effect sizes, Cohen *d*=0.5 for

medium effect sizes, and Cohen *d*=0.8 for large effect sizes) [53]. All analyses were conducted using R software (R Foundation for Statistical Computing) and SAS (version 9.4; SAS Institute).

The focus group discussion data were analyzed by researchers with extensive experience in qualitative analysis using a thematic analysis strategy. The principal investigator checked and verified the codes, and the study team discussed to categorize them into themes.

## Results

### Overview

A total of 13 participants (mean age 76.4, SD 7.9 years) aged between 62 and 86 years took part in the study (Table 4). The participants were primarily women (12/13, 92%), Latino (11/13, 85%), and English speakers (11/13, 85%). Almost all

participants had a high school education or lower (12/13, 92%). All participants (13/13, 100%) met the low-income criteria for eligibility to the state and federal commodity supplemental programs. A total of 8% (1/13) of the participants used an assistance device for balance support, and another participant (1/13, 8%) received a diagnosis of Parkinson disease at the end of the study.

**Table 4.** Study participant characteristics (N=13).

Characteristic	Values
Age (years), mean (SD)	76.4 (7.9)
<b>Sex, n (%)</b>	
Female	12 (92)
Male	1 (8)
<b>Race and ethnicity, n (%)</b>	
White	1 (8)
Latino	11 (85)
Black	1 (8)
<b>Preferred language, n (%)</b>	
English	11 (85)
Spanish	2 (15)
<b>Marital status, n (%)</b>	
Married	3 (23)
Not married	10 (77)
<b>Living arrangement, n (%)</b>	
Lived with someone	3 (23)
Lived alone	10 (77)
<b>Education, n (%)</b>	
Completed elementary school	1 (8)
Completed middle school	1 (8)
Attended high school	3 (23)
Completed high school or GED <sup>a</sup>	7 (54)
Attended college or university	1 (8)
Income—qualified for the Commodity Supplemental Food Program because of low income, n (%)	13 (100)

<sup>a</sup>GED: General Educational Development.

### Feasibility Results

Retention of the participants was very high at the midtest and posttest assessments (Table 5). The completion rate of the data collection protocol for all the healthy aging–related outcomes was 100% (13/13) at baseline and >90% (12/13, 92%) at the midtest and posttest assessments. The attendance to the 24 biweekly group sessions remained high (22.7/24, 94%), ranging from 93% (1.8/2) to 97% (1.9/2) of sessions over the three 4-week periods. Participants were compliant with the incremental weekly home and program exercise goals, with 69.2% (9/13) and 75.0% (9/12) meeting the home and program

goals in the last 4 weeks of the program, respectively. The fidelity of the lesson plan implementation was high, with 92% (33/36) of the sessions delivered as planned. The reasons for not completing the lesson plans were interruption because of technical problems with the television monitor and data collection activities. The level of feasibility of the assessed measures met the criteria of feasibility for the pilot study. There was no report of adverse events related to participation in the study. In all, 8% (1/13) of the participants withdrew from the study after the midtest assessment because of an unrelated, pre-existing health condition.

**Table 5.** Results of study feasibility (N=13).

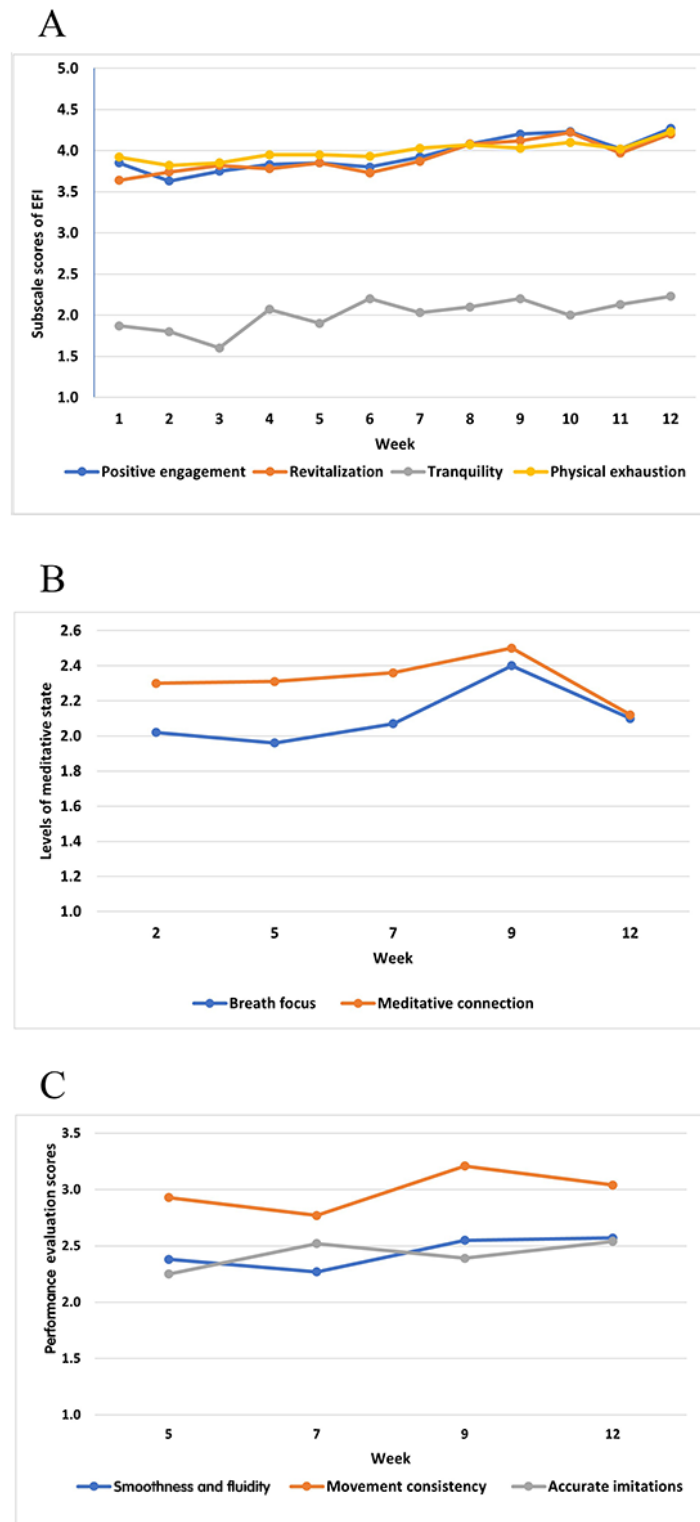
Measurement procedure and time point	Value, n (%)
<b>Participants who took part in the assessments</b>	
Midtest assessment	13 (100)
Posttest assessment	12 (92)
<b>Participants who completed all aging-related outcome measures</b>	
Baseline	13 (100)
Midtest assessment	12 (92)
Posttest assessment	11 (92) <sup>a</sup>
<b>Participants who attended the biweekly group sessions at the senior center</b>	
Weeks 1-4	1.8 (93)
Weeks 5-8	1.8 (93)
Weeks 9-12	1.9 (97)
Weeks 1-12	22.7 (94)
<b>Participants who achieved the weekly home practice goals (<math>\geq 1</math> day per week in weeks 1-4, <math>\geq 2</math> days per week in weeks 5-8, and <math>\geq 3</math> days per week in weeks 9-12)</b>	
Weeks 1-4	9 (75) <sup>a</sup>
Weeks 5-8	7 (58) <sup>a</sup>
Weeks 9-12	9 (69)
<b>Participants who achieved the weekly program exercise goals (center plus home: <math>\geq 3</math> days per week in weeks 1-4, <math>\geq 4</math> days per week in weeks 5-8, and <math>\geq 5</math> days per week in weeks 9-12)</b>	
Weeks 1-4	10 (83) <sup>a</sup>
Weeks 5-8	8 (67) <sup>a</sup>
Weeks 9-12	9 (75) <sup>a</sup>
<b>Lessons in which the lesson activities were delivered as planned</b>	
Weeks 1-4	7 (88) <sup>b</sup>
Weeks 5-8	7 (88) <sup>b</sup>
Weeks 9-12	7 (88) <sup>b</sup>
Weeks 1-12	33 (92) <sup>c</sup>

<sup>a</sup>N=12.<sup>b</sup>N=8.<sup>c</sup>N=36.

Figure 1A shows the modestly increasing trends in tranquility, revitalization, positive engagement, and physical exhaustion at the end of the exercise portion of each session over the 12-week period. The level of participants' meditative state (ie, breath focus and meditative connections) measured by MMI increased gradually from week 2 to week 9 followed by a decrease in

week 12 for both breath focus and meditative connection (Figure 1B). Figure 1C displays the modest upward trends of smooth and fluid movement, movement consistency, and accurate imitations of the Five Animal Frolics routines based on observations of performance by the participants in weeks 5, 7, 9, and 12.

**Figure 1.** (A) Subscale scores (1-5) of the Exercise-Induced Feeling Inventory (EFI): positive engagement, revitalization, tranquility, and physical exhaustion from week 1 to week 12. (B) Levels of meditative state (1-5) over 5 time points for breath focus and meditative connection. (C) Five Animal Play performance evaluation scores (1-5) over 4 time points for smooth and fluid movement, movement consistency, and accurate imitations.



**Acceptability Results**

The level of satisfaction with the FITxOlder program was high, with a majority of participants providing favorable feedback on survey questions addressing the area of program content and support, delivery by CHW, enjoyment and appeal of Five Animal Frolics, opinions on incorporating the background information of qigong and Five Animal Frolics, study burden

and incentives, and safety concerns, with a few exceptions (Multimedia Appendix 2). An overwhelming majority indicated high levels of acceptance of program content (10/11, 91%) and study incentives (11/11, 100%) and enjoyment of Five Animal Frolics (9/10, 90%-10/11, 91%) while expressing little concern for related safety (0%) and learning the cultural background of qigong and Five Animal Frolics (1/11, 9%). The participants



were also very positive about the use of videos, tablets, and SMS text messages for facilitating program delivery and support. The participants also identified areas that could be improved (eg, the delivery of instruction and content by the CHW and long data collection sessions).

Identified themes and examples of corresponding quotes from the participants' responses in the focus group discussion are presented in [Multimedia Appendix 3](#). The most frequently mentioned theme was the *benefits associated with the program* (18 references), and the least frequently mentioned theme was *participation challenges due to family events and doctor's appointments* (3 references). Overall, the FITxOlder program was well received and appreciated. The participants reported health, opportunity for PA, and socialization as motives and time conflicts and family commitments as barriers to program participation. The health benefits associated with the FITxOlder program included improvements in physical and mental health, pain, and sleep quality. The participants also expressed challenges when practicing Five Animal Frolics because of physical limitations or the complexity of the movements and routines but were able to modify (ie, body gestures) to accommodate them. Participants reported experiencing the spirit of the animals and acceptance of being taught qigong information in the instruction. Participants reported a preference for a model proximal to their age (rather than the younger

model) in the Five Animal Frolics video, confusion regarding following the frontal or rear view of the model, and concerns about the timing of the cues in the narration. Finally, the participants reported having shared the exercise in their social circles and expressed a strong desire to continue the exercise in the future.

### Changes in Healthy Aging–Related Outcomes

The responsiveness of the healthy aging–related physiological, cognitive, psychosocial, and behavioral outcomes of the FITxOlder intervention is displayed in [Table 6](#). There were 65% (15/23) of outcome measures that had noticeable differences from baseline to posttest assessment, including (1) changes with small effect sizes in systolic blood pressure, 5-foot fast walk, lean forward reach test, 6-minute walk test, Clock Drawing Test scores, Trail Making Test Part A scores, sleep disturbance, pain severity score, 12-Item Short Form Health Survey (SF-12) mental component scores, sedentary time, and light PA; (2) changes with medium effect sizes in pain interference scores and sleep time; and (3) changes with large effect sizes in the 5-time sit-to-stand test (mean change at posttest assessment =  $-1.62$ ;  $P < .001$ ; Cohen  $d = 0.97$ ) and SF-12 physical component scores (mean change at posttest assessment =  $5.71$ ;  $P = .01$ ; Cohen  $d = 0.88$ ). The changes from baseline to midtest assessment were smaller, as expected, but showed a favorable trend in 43% (10/23) of the outcome measures.

**Table 6.** Health outcomes at baseline, 6 weeks, and 12 weeks.

Variables	Baseline, mean (SD)	Midtest assessment, mean (SD)	Posttest assessment, mean (SD)	Baseline to midtest assessment		Baseline to posttest assessment	
				Change	Cohen <i>d</i>	Change	Cohen <i>d</i>
<b>Anthropometric measures and blood pressure</b>							
Mean SBP <sup>a</sup> (mm Hg)	149.96 (20.80)	139.5 (10.73)	142.75 (15.27)	-10.46	0.62	-7.21	0.39
Mean DBP <sup>b</sup> (mm Hg)	75.27 (11.88)	73.23 (8.92)	76.83 (14.24)	-2.04	0.58	1.56	0.12
BMI (kg/m <sup>2</sup> )	30.13 (3.32)	30.80 (3.68)	29.99 (4.35)	0.66 <sup>c</sup>	0.19	-0.15	0.04
Body fat percentage, %	41.22 (4.81)	41.57 (4.7)	41.46 (5.6)	0.35	0.07	0.24	0.03
<b>Physical function</b>							
5XSTS <sup>d</sup> (seconds)	12.68 (2.45)	11.93 (3.59)	10.74 (2.21)	-0.75	0.24	-1.62 <sup>c</sup>	0.97
Grip strength (dominant hand; kg)	19.10 (6.82)	18.45 (5.8)	19.61 (5.96)	-0.65	0.23	0.11	0.05
FWT <sup>e</sup> (seconds)	16.00 (5.03)	18.82 (9.82)	15.06 (4.73)	2.82	0.36	-0.94	0.20
Reach (cm)	25.35 (6.55)	22.75 (9.24)	27.42 (4.57)	2.6	0.33	2.06	0.36
6MWT <sup>f</sup> (m)	383.28 (116.67)	392.56 (130.32)	412.02 (118.72)	9.28	0.08	28.73	0.24
<b>Cognitive function</b>							
Clock Drawing Test score	7.46 (1.98)	8.62 (1.26)	8.08 (1.68)	1.15 <sup>g</sup>	0.69	0.62	0.24
TMT <sup>h</sup> Part A (seconds)	73.14 (68.09)	62.89 (61.12)	58.23 (54.19)	-10.26	0.16	-14.91	0.24
TMT Part B (seconds)	161.72 (107.66)	166.89 (161.33)	149.87 (166.44)	5.17	0.04	-11.84	0.09
<b>Psychosocial and behavioral measures</b>							
Sleep disturbance score	23.38 (2.53)	23.23 (2.59)	22.67 (2.74)	-0.15	0.06	-0.72	0.27
Pain severity score	2.83 (1.92)	2.52 (2.85)	2.35 (1.87)	-0.31	0.13	-0.47	0.25
Pain interference score	2.89 (2.64)	2.23 (2.37)	1.47 (1.72)	-0.66	0.26	1.42	0.64
Basic ADL <sup>i</sup> score	10.69 (3.28)	11.31 (1.03)	10.67 (3.42)	0.62	0.25	0.03	0.01
Intermediate ADL score	18.62 (5.42)	18.92 (3.97)	18.92 (3.70)	0.31	0.07	0.31	0.06
FSQ <sup>j</sup> total score	29.31 (8.02)	30.23 (4.21)	29.58 (6.4)	0.92	0.14	0.28	0.04
SF-12-PCS <sup>k</sup> score	39.15 (9)	42.78 (9.75)	46.43 (8.6)	3.64	0.39	5.71 <sup>g</sup>	0.88
SF-12-MCS <sup>l</sup> score	53.80 (9.87)	53.87 (12.03)	50.57 (11.06)	0.07	0.01	-3.23	0.30
Sleep (hours per day)	6.29 (1.36)	6.30 (1.49)	6.83 (2.49)	0	0	0.53	0.52
Sedentary time (hours per day)	13.92 (2.28)	13.85 (2.23)	12.92 (3.35)	-0.07	0.03	-0.99	0.35
LPA <sup>m</sup> (hours per day)	2.78 (1.39)	2.89 (1.04)	3.26 (1.12)	0.1	0.1	0.47	0.37

<sup>a</sup>SBP: systolic blood pressure.<sup>b</sup>DBP: diastolic blood pressure.<sup>c</sup>*P* < .01.<sup>d</sup>5XSTS: 5-time sit-to-stand test.<sup>e</sup>FWT: 50-foot fast walk test.<sup>f</sup>6MWT: 6-minute walk test.<sup>g</sup>*P* = .01.<sup>h</sup>TMT: Trail Making Test.<sup>i</sup>ADL: activities of daily living.<sup>j</sup>FSQ: Functional Status Questionnaire.<sup>k</sup>SF-12-PCS: 12-Item Short Form Health Survey physical component score.<sup>l</sup>SF-12-MCS: 12-Item Short Form Health Survey mental component score.<sup>m</sup>LPA: light physical activity.

## Discussion

### Principal Findings

The results of the study support the feasibility and acceptability of the refined FITxOlder, a community-based mind-body exercise program based on the Five Animal Frolics delivered by a CHW with technology-facilitated support and tailored for low-income community-dwelling Latino older adults. Strong feasibility was demonstrated by participant retention and program attendance; compliance with exercise goals was indicative of the increased appeal of complementary and alternative medicine and mind-body exercise [54] for healthy aging in older adults who are seeking a socially, culturally, and geragogically tailored program [33,55,56]. The trending health benefits in physical function, cognition, and psychosocial well-being as well as objective measures of PA and sleep are essential for healthy aging and independent living [2,3].

### Feasibility and Acceptability

The high level of engagement and retention in this study can be partially attributed to the engagement of older adults in developing the PA program [57]. FITxOlder was based on lessons learned from a previous pilot study that used an iterative process to gather input from a work group of Latino older adults in developing and tailoring the delivery of the FITxOlder program as well as the Five Animal Frolics routines [38]. For example, participants were appreciative of the group format that offered opportunities for socialization and peer support in FITxOlder [58,59] and the program delivery by a bilingual CHW who provided linguistically and socially sensitive support to engage the participants [60]. The findings also reflect older adults' interest in improving quality of life and independent living beyond health indexes [61]. In addition, FITxOlder was designed as a safe and low to moderate intensity mind-body exercise, which was favored by older adults with considerable health conditions that hindered them from engaging in intensive exercise programs [62]. However, feedback from the participants revealed needs for improvement in the quality and design of the exercise video and instruction delivery. Finally, the use of the exercise video modeled by the CHW and of SMS text message reminders facilitated the participation in home practice and the achievement of weekly exercise goals. The overall adherence to the prescribed exercise frequency in the study was similar to the adherence rate in published exercise studies on older adults [63].

The upward trend in tranquility, revitalization, and positive engagement scores of the Exercise-Induced Feeling Inventory subscales reflected positive psychological and emotional reactions in the participants that have been associated with increased PA level as well as exercise adherence in previously published studies on adults [64,65] and older adults [66,67]. Participants in a 10-week yoga intervention group also demonstrated increased tranquility, revitalization, and positive engagement [68]. Furthermore, it is worth pointing out that the scores on the physical exhaustion subscale at each group session consistently exceeded the level reported by older adults who completed a 30-minute moderate-intensity aerobic exercise routine ( $\geq 64\%$  of maximum heart rate) [66].

The gradual increases in the scores of breath focus and meditative connection on the MMI and in the observed movement smoothness and fluidity and accuracy of movement imitation were consistent with the progressive goals in the 3 steps of the instructions (ie, the participants' meditative state and movement proficiency improved with in-class instruction and practice during the first 9 weeks [43]). The decline in the scores in week 12 was the result of introducing the advanced version of the Five Animal Frolics that started in week 10. Although it is normal to experience reduced mastery when a more challenging routine is introduced, the decline in the scores indicated that the participants were inadequately prepared to learn the advanced version, and thus their experience of mastery was adversely affected. However, the overall changes in scores suggest that the participants were responsive to the stepped instruction and support the validity of the MMI and performance observation.

The refined FITxOlder was highly acceptable to the study participants, as demonstrated by the high retention, high program attendance, and compliance with weekly exercise goals by FITxOlder participants. Our results were consistent with the findings of a 2021 systematic review that reported high attendance, with 73% to 95% participation in supervised group sessions and adherence to home self-practice (63%-80%) in women taking part in qigong interventions [69]. However, past healthy aging studies using traditional Western exercise or tai chi have reported low retention and low program participation because of dissatisfaction with the program delivery format or the program not meeting the participants' needs (disease focus vs holistic health) [70,71]. For example, HAPPY, a community-based healthy aging program of biweekly group sessions delivered by a peer coach using traditional Western exercise, significantly improved physical and cognitive functions but only retained 66.6% of older adult participants at the 3-month follow-up [72]. Furthermore, results from our quantitative and qualitative analysis indicated that the Five Animal Frolics being easy to learn, suitable for home practice, and safe. The acceptance of using a tablet for FITxOlder practice suggests that Latino older adults can benefit from using digital technologies provided by the study for health promotion practices [73,74]. A previous study reported that older participants were able to practice qigong at home using a study-provided video [75]. FITxOlder participants also expressed the importance of SMS text message reminders for keeping them on track to meet weekly exercise goals. Of note, the percentage of participants who achieved the weekly home practice goals was not optimal and was lower than the rate of attendance to group sessions. This suggests that the at-home practice recommendation did help increase exercise engagement to some extent; it also suggests the importance of a supportive learning environment to promote exercise engagement. The completion rate (11/12, 91.7%-13/13, 100%) of data collection was higher than those reported in past qigong studies (67%) [69], and most participants (9/10, 90%) did not consider the data collection burdensome. Finally, this study offered important insight on the acceptance of introducing background information on qigong and Five Animal Frolics in a predominantly Latino sample. In the future, it is important to further explore the best approach for the integration of the multidimensions of qigong

into the holistic health framework for effective dissemination to non-Asian population groups [76,77].

### Healthy Aging–Related Outcomes

Despite the light to moderate intensity of the Five Animal Frolics, the responsiveness in the healthy aging–related outcomes was promising and showed an effect size consistent with published randomized controlled trials in older adults with longer intervention durations using traditional Western exercises [78], tai chi [79,80], or Five Animal Frolics [81,82]. Of particular interest were the changes in the 5-time sit-to-stand and SF-12 physical component scores, which were reflective of improvements in physical functions for independent living among older adults [83,84]. To be noted, there was also an improvement in cognitive function, another key factor for healthy aging and independent living [85,86]. By contrast, the decline in SF-12 mental component scores from midtest to posttest assessments was counterintuitive to the reported socialization benefit from participating in the FITxOlder program. We speculate that worsening health conditions (ie, diagnosis of mental disorder and surgery) unrelated to the study and family hardships in some of the participants in the last 4 weeks of the intervention might have contributed to the decline in mental health. Finally, the slightly increased BMI from baseline to midtest assessment was unexpected. On the basis of the report of a review, overall studies directly examining the effects of mind-body exercise on body weight are limited [87]. A recent study randomized 543 participants into a tai chi group, a conventional exercise group, and a control group and reported reduced body weight and waist circumference in both intervention groups but not in the control group at weeks 12 and 38 [88]. The midtest assessment in our study at 6 weeks after the intervention may not have been long enough to show exercise effects. Particularly, our study did not consider other factors such as diet intake, which are also important for body weight and BMI.

There was indirect evidence that the FITxOlder program helped participants meet PA recommendations [89] as participants practiced  $\geq 30$  minutes per day of Five Animal Frolics on most days of the week during the last month of the intervention. The increase in PA was corroborated by objectively measured changes in time spent in light PA (+28.2 minutes per day) as well as sedentary time (–59.4 minutes per day). Of note, the

increase in light PA is particularly important as Five Animal Frolics consists of slow and smooth movements that will be recorded as light rather than moderate PA by accelerometry. The increase in sleep duration (+38.3 minutes per day) was another benefit for maintaining physical and cognitive functioning in older adults [90].

### Limitations

There are several limitations to this study. The changes observed in the study should be interpreted with caution because of the short intervention duration and use of a single-arm design. Previous studies have reported significant improvements in the outcomes of older adults using interventions lasting  $\geq 6$  months [91]. The single-arm design also did not allow for the testing of a placebo effect associated with within-study socialization and attention received by the study staff [92]. The efficacy of qigong for healthy aging should be tested using a rigorous study design, a larger sample size, and long-term follow-up. In addition, the improvement in physical and cognitive function outcomes may be due to a learning effect caused by repeated testing. Not all the participants in this study were Latino. However, all participants (13/13, 100%) were older adults from the same low-income community, and most of them (11/13, 85%) were Latino. Therefore, our findings could still be applicable to our target population of low-income community-dwelling older Latino individuals. Finally, most of the study participants (12/13, 92%) were women. The generalizability of the findings to men is limited. Strategies for recruiting male participants to such studies need to be explored.

### Conclusions

Healthy aging research in underserved minority populations remains limited [93]. The FITxOlder program demonstrated high feasibility and acceptability for both in-person delivery and practice at home among low-income Latino older adults. The trending health benefits of this short-term intervention suggest that the FITxOlder program might be a promising approach to promote PA engagement in underserved older populations for various health improvements (eg, physical and cognitive function and psychosocial well-being), which are essential for the promotion of healthy aging and independent living. Future research with larger samples and longer interventions is warranted to assess the health benefits of the FITxOlder program and its suitability.

### Acknowledgments

The authors wish to express their heartfelt appreciation to the study participants for dedicating their time to the study and to Ms Mary Catherine Fernandez of the Salvation Am Peacock Center, San Antonio, Texas, for coordinating and supporting the study. They also thank Ms Rosa Yin for editing the manuscript. Finally, gratitude goes to the undergraduate internship students and volunteer students who assisted with data collection and evaluation.

### Conflicts of Interest

None declared.

Multimedia Appendix 1

Biweekly group session format.

[[DOCX File , 27 KB - formative\\_v6i9e40046\\_app1.docx](#) ]

## Multimedia Appendix 2

The 12-week older adult exercise study poststudy survey.

[[DOCX File , 42 KB - formative\\_v6i9e40046\\_app2.docx](#) ]

## Multimedia Appendix 3

Themes from the focus group discussion.

[[DOCX File , 31 KB - formative\\_v6i9e40046\\_app3.docx](#) ]

**References**

1. Healthy ageing and functional ability. World Health Organization. 2020 Oct 26. URL: <https://www.who.int/news-room/questions-and-answers/item/healthy-ageing-and-functional-ability> [accessed 2022-05-20]
2. Fausset CB, Kelly AJ, Rogers WA, Fisk AD. Challenges to aging in place: understanding home maintenance difficulties. *J Hous Elderly* 2011;25(2):125-141 [FREE Full text] [doi: [10.1080/02763893.2011.571105](https://doi.org/10.1080/02763893.2011.571105)] [Medline: [22072843](https://pubmed.ncbi.nlm.nih.gov/22072843/)]
3. Passel JS, Cohn D. U.S. Population Projections: 2005-2050. Pew Research Center. 2008 Feb 11. URL: <https://www.pewresearch.org/hispanic/2008/02/11/us-population-projections-2005-2050/> [accessed 2022-05-20]
4. Tovar JJ, Angel RJ, Eschbach K, Espino DV, Markides KS. Hispanic established populations for the epidemiologic studies of the elderly: selected longitudinal findings. *Aging Health* 2007 Jun;3(3):325-331. [doi: [10.2217/1745509x.3.3.325](https://doi.org/10.2217/1745509x.3.3.325)]
5. Jimenez DE, Martinez Garza D, Cárdenas V, Marquine M. Older Latino mental health: a complicated picture. *Innov Aging* 2020 Aug 18;4(5):igaa033 [FREE Full text] [doi: [10.1093/geroni/igaa033](https://doi.org/10.1093/geroni/igaa033)] [Medline: [32964142](https://pubmed.ncbi.nlm.nih.gov/32964142/)]
6. Usher T, Buta B, Thorpe RJ, Huang J, Samuel LJ, Kasper JD, et al. Dissecting the racial/ethnic disparity in frailty in a nationally representative cohort study with respect to health, income, and measurement. *J Gerontol A Biol Sci Med Sci* 2021 Jan 01;76(1):69-76 [FREE Full text] [doi: [10.1093/gerona/glaa061](https://doi.org/10.1093/gerona/glaa061)] [Medline: [32147727](https://pubmed.ncbi.nlm.nih.gov/32147727/)]
7. Hayward MD, Hummer RA, Chiu CT, González-González C, Wong R. Does the Hispanic paradox in U.S. adult mortality extend to disability? *Popul Res Policy Rev* 2014 Feb 01;33(1):81-96 [FREE Full text] [doi: [10.1007/s11113-013-9312-7](https://doi.org/10.1007/s11113-013-9312-7)] [Medline: [25821283](https://pubmed.ncbi.nlm.nih.gov/25821283/)]
8. Prickett KC, Angel JL. Transitions in living arrangements among older Mexican Americans. *Res Aging* 2017 Mar;39(3):396-417 [FREE Full text] [doi: [10.1177/0164027515613141](https://doi.org/10.1177/0164027515613141)] [Medline: [26658456](https://pubmed.ncbi.nlm.nih.gov/26658456/)]
9. Kim J, Angel JL, Rote SM. A longitudinal study of cognitive and instrumental activities of daily living disablement among the oldest Mexican Americans. *J Aging Health* 2022 Mar;34(2):196-205 [FREE Full text] [doi: [10.1177/08982643211037512](https://doi.org/10.1177/08982643211037512)] [Medline: [34388944](https://pubmed.ncbi.nlm.nih.gov/34388944/)]
10. Kojima G. Frailty as a predictor of disabilities among community-dwelling older people: a systematic review and meta-analysis. *Disabil Rehabil* 2017 Sep;39(19):1897-1908. [doi: [10.1080/09638288.2016.1212282](https://doi.org/10.1080/09638288.2016.1212282)] [Medline: [27558741](https://pubmed.ncbi.nlm.nih.gov/27558741/)]
11. Garcia MA, Downer B, Crowe M, Markides KS. Aging and disability among Hispanics in the United States: current knowledge and future directions. *Innov Aging* 2017 Sep;1(2):igx020 [FREE Full text] [doi: [10.1093/geroni/igx020](https://doi.org/10.1093/geroni/igx020)] [Medline: [30480115](https://pubmed.ncbi.nlm.nih.gov/30480115/)]
12. National Center for Complementary and Alternative Medicine - Third Strategic Plan: 2011-2015. Exploring the Science of Complementary and Alternative Medicine. U.S. Department of Health and Human Services, National Institutes of Health. 2011. URL: <https://files.nccih.nih.gov/exploring-the-science-of-complementary-and-alternative-medicine-strategic-plan-2011-2015.pdf> [accessed 2022-05-20]
13. Larkey L, Jahnke R, Etnier J, Gonzalez J. Meditative movement as a category of exercise: implications for research. *J Phys Act Health* 2009 Mar;6(2):230-238. [doi: [10.1123/jpah.6.2.230](https://doi.org/10.1123/jpah.6.2.230)] [Medline: [19420401](https://pubmed.ncbi.nlm.nih.gov/19420401/)]
14. Pölonen P, Lappi O, Tervaniemi M. Effect of meditative movement on affect and flow in qigong practitioners. *Front Psychol* 2019 Oct 22;10:2375 [FREE Full text] [doi: [10.3389/fpsyg.2019.02375](https://doi.org/10.3389/fpsyg.2019.02375)] [Medline: [31695654](https://pubmed.ncbi.nlm.nih.gov/31695654/)]
15. Clarke TC, Black LI, Stussman BJ, Barnes PM, Nahin RL. Trends in the use of complementary health approaches among adults: United States, 2002-2012. *Natl Health Stat Report* 2015 Feb 10(79):1-16 [FREE Full text] [Medline: [25671660](https://pubmed.ncbi.nlm.nih.gov/25671660/)]
16. Lauche R, Wayne PM, Dobos G, Cramer H. Prevalence, patterns, and predictors of T'ai chi and qigong use in the United States: results of a nationally representative survey. *J Altern Complement Med* 2016 Apr;22(4):336-342. [doi: [10.1089/acm.2015.0356](https://doi.org/10.1089/acm.2015.0356)] [Medline: [26981616](https://pubmed.ncbi.nlm.nih.gov/26981616/)]
17. Sharp D, Lorenc A, Morris R, Feder G, Little P, Hollinghurst S, et al. Complementary medicine use, views, and experiences: a national survey in England. *BJGP Open* 2018 Dec;2(4):bjgpopen18X101614 [FREE Full text] [doi: [10.3399/bjgpopen18X101614](https://doi.org/10.3399/bjgpopen18X101614)] [Medline: [30723800](https://pubmed.ncbi.nlm.nih.gov/30723800/)]
18. Jahnke R, Larkey L, Rogers C, Etnier J, Lin F. A comprehensive review of health benefits of qigong and tai chi. *Am J Health Promot* 2010;24(6):e1-25 [FREE Full text] [doi: [10.4278/ajhp.081013-LIT-248](https://doi.org/10.4278/ajhp.081013-LIT-248)] [Medline: [20594090](https://pubmed.ncbi.nlm.nih.gov/20594090/)]
19. Tsang H, Chan CL, Chen K, Cho WC, Lee MS. Scientific basis of mind-body interventions. *Evid Based Complement Alternat Med* 2014;2014:706892 [FREE Full text] [doi: [10.1155/2014/706892](https://doi.org/10.1155/2014/706892)] [Medline: [24734112](https://pubmed.ncbi.nlm.nih.gov/24734112/)]
20. Wang YT, Huang G, Duke G, Yang Y. Tai chi, yoga, and qigong as mind-body exercises. *Evid Based Complement Alternat Med* 2017;2017:8763915 [FREE Full text] [doi: [10.1155/2017/8763915](https://doi.org/10.1155/2017/8763915)] [Medline: [28154608](https://pubmed.ncbi.nlm.nih.gov/28154608/)]

21. Morone NE, Greco CM. Mind-body interventions for chronic pain in older adults: a structured review. *Pain Med* 2007;8(4):359-375. [doi: [10.1111/j.1526-4637.2007.00312.x](https://doi.org/10.1111/j.1526-4637.2007.00312.x)] [Medline: [17610459](https://pubmed.ncbi.nlm.nih.gov/17610459/)]
22. Fogaça LZ, Portella CF, Ghelman R, Abdala CV, Schweitzer MC. Mind-body therapies from traditional Chinese medicine: evidence map. *Front Public Health* 2021 Dec 10;9:659075 [FREE Full text] [doi: [10.3389/fpubh.2021.659075](https://doi.org/10.3389/fpubh.2021.659075)] [Medline: [34988045](https://pubmed.ncbi.nlm.nih.gov/34988045/)]
23. Payne P, Crane-Godreau MA. Meditative movement for depression and anxiety. *Front Psychiatry* 2013 Jul 24;4:71 [FREE Full text] [doi: [10.3389/fpsy.2013.00071](https://doi.org/10.3389/fpsy.2013.00071)] [Medline: [23898306](https://pubmed.ncbi.nlm.nih.gov/23898306/)]
24. Birdee GS, Wayne PM, Davis RB, Phillips RS, Yeh GY. T'ai chi and qigong for health: patterns of use in the United States. *J Altern Complement Med* 2009 Sep;15(9):969-973 [FREE Full text] [doi: [10.1089/acm.2009.0174](https://doi.org/10.1089/acm.2009.0174)] [Medline: [19757974](https://pubmed.ncbi.nlm.nih.gov/19757974/)]
25. Nayak HD, Patel NK, Wood R, Dufault V, Guidotti N. A study to identify the benefits, barriers, and cues to participating in a yoga program among community dwelling older adults. *J Yoga Phys Ther* 2015 Jan 1;05(01):1. [doi: [10.4172/2157-7595.1000178](https://doi.org/10.4172/2157-7595.1000178)]
26. Balaneskovic S. Hua Tuo's Wu Qin Xi (Five Animal Frolics) movements and the logic behind it. *Chin Med Cult* 2018;1(3):127-134. [doi: [10.4103/cmac.cmac\\_32\\_18](https://doi.org/10.4103/cmac.cmac_32_18)]
27. Chinese Health Qigong Association. *Wu Qin Xi: Five-Animal Qigong Exercises (Chinese Health Qigong)*. Beijing, China: Foreign Language Press; 2008.
28. Rogers CE, Larkey LK, Keller C. A review of clinical trials of tai chi and qigong in older adults. *West J Nurs Res* 2009 Mar;31(2):245-279 [FREE Full text] [doi: [10.1177/0193945908327529](https://doi.org/10.1177/0193945908327529)] [Medline: [19179544](https://pubmed.ncbi.nlm.nih.gov/19179544/)]
29. Nguyen SA, LAVretsky H. Emerging complementary and integrative therapies for geriatric mental health. *Curr Treat Options Psychiatry* 2020;7(4):447-470 [FREE Full text] [doi: [10.1007/s40501-020-00229-5](https://doi.org/10.1007/s40501-020-00229-5)] [Medline: [32904865](https://pubmed.ncbi.nlm.nih.gov/32904865/)]
30. Rogers C, Keller C, Larkey LK. Perceived benefits of meditative movement in older adults. *Geriatr Nurs* 2010;31(1):37-51. [doi: [10.1016/j.gerinurse.2009.10.002](https://doi.org/10.1016/j.gerinurse.2009.10.002)] [Medline: [20159352](https://pubmed.ncbi.nlm.nih.gov/20159352/)]
31. Farhang M, Miranda-Castillo C, Rubio M, Furtado G. Impact of mind-body interventions in older adults with mild cognitive impairment: a systematic review. *Int Psychogeriatr* 2019 May;31(5):643-666. [doi: [10.1017/S1041610218002302](https://doi.org/10.1017/S1041610218002302)] [Medline: [30712518](https://pubmed.ncbi.nlm.nih.gov/30712518/)]
32. Meikis L, Wicker P, Donath L. Effects of pilates training on physiological and psychological health parameters in healthy older adults and in older adults with clinical conditions over 55 years: a meta-analytical review. *Front Neurol* 2021 Oct 25;12:724218 [FREE Full text] [doi: [10.3389/fneur.2021.724218](https://doi.org/10.3389/fneur.2021.724218)] [Medline: [34759879](https://pubmed.ncbi.nlm.nih.gov/34759879/)]
33. Zubala A, MacGillivray S, Frost H, Kroll T, Skelton DA, Gavine A, et al. Promotion of physical activity interventions for community dwelling older adults: a systematic review of reviews. *PLoS One* 2017 Jul 10;12(7):e0180902 [FREE Full text] [doi: [10.1371/journal.pone.0180902](https://doi.org/10.1371/journal.pone.0180902)] [Medline: [28700754](https://pubmed.ncbi.nlm.nih.gov/28700754/)]
34. Bernal G, Jiménez-Chafey MI, Domenech Rodríguez MM. Cultural adaptation of treatments: a resource for considering culture in evidence-based practice. *Prof Psychol Res Pr* 2009 Aug;40(4):361-368. [doi: [10.1037/a0016401](https://doi.org/10.1037/a0016401)]
35. Bao GC. The idealist and pragmatist view of qi in tai chi and qigong: a narrative commentary and review. *J Integr Med* 2020 Sep;18(5):363-368. [doi: [10.1016/j.joim.2020.06.004](https://doi.org/10.1016/j.joim.2020.06.004)] [Medline: [32636157](https://pubmed.ncbi.nlm.nih.gov/32636157/)]
36. Patwardhan B, Mutalik G, Tillu G. Concepts of health and disease. In: Patwardhan B, Mutalik G, Tillu G, editors. *Integrative Approaches for Health: Biomedical Research, Ayurveda and Yoga*. Boston, MA, USA: Academic Press; 2015:53-78.
37. Onken LS, Carroll KM, Shoham V, Cuthbert BN, Riddle M. Reenvisioning clinical science: unifying the discipline to improve the public health. *Clin Psychol Sci* 2014 Jan 01;2(1):22-34 [FREE Full text] [doi: [10.1177/2167702613497932](https://doi.org/10.1177/2167702613497932)] [Medline: [25821658](https://pubmed.ncbi.nlm.nih.gov/25821658/)]
38. Yin Z, Martinez CE, Li S, Martinez M, Peng K, Land WM, et al. Adapting Chinese qigong mind-body exercise for healthy aging in older community-dwelling low-income Latino adults: pilot feasibility study. *JMIR Aging* 2021 Nov 01;4(4):e29188 [FREE Full text] [doi: [10.2196/29188](https://doi.org/10.2196/29188)] [Medline: [34723824](https://pubmed.ncbi.nlm.nih.gov/34723824/)]
39. Fisher WA, Fisher JD, Harman J. The information-motivation-behavioral skills model: a general social psychological approach to understanding and promoting health behavior. In: Suls J, Wallston KA, editors. *Social Psychological Foundations of Health and Illness*. Hoboken, NJ, USA: Blackwell Publishing; 2003:82-106.
40. John MT. Geragogy: a theory for teaching the elderly. *Activities Adaptation Aging* 1988;11(3-4):xi-156. [doi: [10.1300/J016v11n03\\_a](https://doi.org/10.1300/J016v11n03_a)]
41. Davis LA, Chesbro SB. Integrating health promotion, patient education, and adult education principles with the older adult: a perspective for rehabilitation professionals. *J Allied Health* 2003;32(2):106-109. [Medline: [12801022](https://pubmed.ncbi.nlm.nih.gov/12801022/)]
42. Meyer SL. Andragogy and the aging adult learner. *Educ Gerontol* 1977;2(2):115-122. [doi: [10.1080/03601277700202020](https://doi.org/10.1080/03601277700202020)]
43. Larkey L, Szalacha L, Rogers C, Jahnke R, Ainsworth B. Measurement pilot study of the meditative movement inventory (MMI). *J Nurs Meas* 2012;20(3):230-243. [doi: [10.1891/1061-3749.20.3.230](https://doi.org/10.1891/1061-3749.20.3.230)] [Medline: [23362559](https://pubmed.ncbi.nlm.nih.gov/23362559/)]
44. Simmonds MJ, Olson SL, Jones S, Hussein T, Lee CE, Novy D, et al. Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. *Spine (Phila Pa 1976)* 1998 Nov 15;23(22):2412-2421. [doi: [10.1097/00007632-199811150-00011](https://doi.org/10.1097/00007632-199811150-00011)] [Medline: [9836355](https://pubmed.ncbi.nlm.nih.gov/9836355/)]
45. Freedman M, Leach L, Kaplan E, Wincour G, Shulman K, Delis DC. *Clock Drawing: A Neuropsychological Analysis*. Oxford, UK: Oxford University Press; 1994.

46. Halstead WC. *Brain and Intelligence: A Quantitative Study of the Frontal Lobes*. Chicago, IL, USA: University of Chicago Press; 1947.
47. Ware Jr JE, Kosinski M, Turner-Bowker DM, Sundaram M, Gandek B, Maruish ME. *User's Manual for the SF-12v2 Health Survey*. 2nd edition. Johnston, RI, USA: QualityMetric; 2009.
48. Poquet N, Lin C. The brief pain inventory (BPI). *J Physiother* 2016 Jan;62(1):52 [FREE Full text] [doi: [10.1016/j.jphys.2015.07.001](https://doi.org/10.1016/j.jphys.2015.07.001)] [Medline: [26303366](https://pubmed.ncbi.nlm.nih.gov/26303366/)]
49. 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](https://doi.org/10.1080/15402002.2012.636266)] [Medline: [22250775](https://pubmed.ncbi.nlm.nih.gov/22250775/)]
50. Jette AM, Davies AR, Cleary PD, Calkins DR, Rubenstein LV, Fink A, et al. The Functional Status Questionnaire: reliability and validity when used in primary care. *J Gen Intern Med* 1986;1(3):143-149. [doi: [10.1007/BF02602324](https://doi.org/10.1007/BF02602324)] [Medline: [3772582](https://pubmed.ncbi.nlm.nih.gov/3772582/)]
51. Migueles JH, Rowlands AV, Huber F, Sabia S, van Hees VT. GGIR: a research community-driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. *J Meas Phys Behav* 2019 Sep 23;2(3):188-196. [doi: [10.1123/jmpb.2018-0063](https://doi.org/10.1123/jmpb.2018-0063)]
52. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. *Nurs Health Sci* 2013 Sep;15(3):398-405. [doi: [10.1111/nhs.12048](https://doi.org/10.1111/nhs.12048)] [Medline: [23480423](https://pubmed.ncbi.nlm.nih.gov/23480423/)]
53. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edition. New York, NY, USA: Routledge; 1988.
54. Cartwright T. 'Getting on with life': the experiences of older people using complementary health care. *Soc Sci Med* 2007 Apr;64(8):1692-1703. [doi: [10.1016/j.socscimed.2006.11.034](https://doi.org/10.1016/j.socscimed.2006.11.034)] [Medline: [17270332](https://pubmed.ncbi.nlm.nih.gov/17270332/)]
55. Purdie N, Boulton-Lewis G. The learning needs of older adults. *Educ Gerontol* 2003 Feb;29(2):129-149. [doi: [10.1080/713844281](https://doi.org/10.1080/713844281)]
56. Pasick RJ, D'onofrio CN, Otero-Sabogal R. Similarities and differences across cultures: questions to inform a third generation for health promotion research. *Health Educ Q* 1996;23(1\_suppl):142-161. [doi: [10.1177/109019819602301s11](https://doi.org/10.1177/109019819602301s11)]
57. Boulton ER, Horne M, Todd C. Involving older adults in developing physical activity interventions to promote engagement: a literature review. *J Population Ageing* 2019 Jun 24;13(3):325-345. [doi: [10.1007/s12062-019-09247-5](https://doi.org/10.1007/s12062-019-09247-5)]
58. Belza B, Walwick J, Shiu-Thornton S, Schwartz S, Taylor M, LoGerfo J. Older adult perspectives on physical activity and exercise: voices from multiple cultures. *Prev Chronic Dis* 2004 Oct;1(4):A09 [FREE Full text] [Medline: [15670441](https://pubmed.ncbi.nlm.nih.gov/15670441/)]
59. Ickes MJ, Sharma M. A systematic review of physical activity interventions in Hispanic adults. *J Environ Public Health* 2012;2012:156435 [FREE Full text] [doi: [10.1155/2012/156435](https://doi.org/10.1155/2012/156435)] [Medline: [22496702](https://pubmed.ncbi.nlm.nih.gov/22496702/)]
60. Kennedy MA, Hatchell KE, DiMilia PR, Kelly SM, Blunt HB, Bagley PJ, et al. Community health worker interventions for older adults with complex health needs: a systematic review. *J Am Geriatr Soc* 2021 Jun;69(6):1670-1682 [FREE Full text] [doi: [10.1111/jgs.17078](https://doi.org/10.1111/jgs.17078)] [Medline: [33738803](https://pubmed.ncbi.nlm.nih.gov/33738803/)]
61. Borglin G, Edberg AK, Rahm Hallberg I. The experience of quality of life among older people. *J Aging Stud* 2005 May;19(2):201-220 [FREE Full text] [doi: [10.1016/j.jaging.2004.04.001](https://doi.org/10.1016/j.jaging.2004.04.001)]
62. Roberts CE, Phillips LH, Cooper CL, Gray S, Allan JL. Effect of different types of physical activity on activities of daily living in older adults: systematic review and meta-analysis. *J Aging Phys Act* 2017 Oct 01;25(4):653-670. [doi: [10.1123/japa.2016-0201](https://doi.org/10.1123/japa.2016-0201)] [Medline: [28181837](https://pubmed.ncbi.nlm.nih.gov/28181837/)]
63. Killingback C, Tsofliou F, Clark C. Older people's adherence to community-based group exercise programmes: a multiple-case study. *BMC Public Health* 2017 Jan 25;17(1):115 [FREE Full text] [doi: [10.1186/s12889-017-4049-6](https://doi.org/10.1186/s12889-017-4049-6)] [Medline: [28122532](https://pubmed.ncbi.nlm.nih.gov/28122532/)]
64. Annesi JJ. Relationship between changes in acute exercise-induced feeling states, self-motivation, and adults' adherence to moderate aerobic exercise. *Percept Mot Skills* 2002 Apr;94(2):425-439. [doi: [10.2466/pms.2002.94.2.425](https://doi.org/10.2466/pms.2002.94.2.425)] [Medline: [12027334](https://pubmed.ncbi.nlm.nih.gov/12027334/)]
65. Lacharité-Lemieux M, Brunelle JP, Dionne IJ. Adherence to exercise and affective responses: comparison between outdoor and indoor training. *Menopause* 2015 Jul;22(7):731-740. [doi: [10.1097/GME.0000000000000366](https://doi.org/10.1097/GME.0000000000000366)] [Medline: [25423324](https://pubmed.ncbi.nlm.nih.gov/25423324/)]
66. Brunet J, Guérin E, Speranzini N. An examination of exercise-induced feeling states and their association with future participation in physical activity among older adults. *J Aging Phys Act* 2018 Jan 01;26(1):52-60. [doi: [10.1123/japa.2016-0342](https://doi.org/10.1123/japa.2016-0342)] [Medline: [28422538](https://pubmed.ncbi.nlm.nih.gov/28422538/)]
67. McAuley E, Blissmer B, Katula J, Duncan TE. Exercise environment, self-efficacy, and affective responses to acute exercise in older adults. *Psychol Health* 2000 May;15(3):341-355. [doi: [10.1080/08870440008401997](https://doi.org/10.1080/08870440008401997)]
68. Bryan S, Pinto Zipp G, Parasher R. The effects of yoga on psychosocial variables and exercise adherence: a randomized, controlled pilot study. *Altern Ther Health Med* 2012;18(5):50-59. [Medline: [22894891](https://pubmed.ncbi.nlm.nih.gov/22894891/)]
69. Leung KW, Yang YJ, Hui SS, Woo J. Mind-body health benefits of traditional Chinese qigong on women: a systematic review of randomized controlled trials. *Evid Based Complement Alternat Med* 2021 Sep 14;2021:7443498 [FREE Full text] [doi: [10.1155/2021/7443498](https://doi.org/10.1155/2021/7443498)] [Medline: [34567220](https://pubmed.ncbi.nlm.nih.gov/34567220/)]
70. Chen K. Methodological challenges and research design in research study of qigong therapies. In: McCarthy PW, Langweiler MJ, editors. *Methodologies for Effectively Assessing Complementary and Alternative Medicine (CAM): Research Tools and Techniques*. Beijing, China: Singing Dragon; 2015:228-248.

71. Dong X, Bergren S. Qigong among older adults: a global review. *Clin Res Trial* 2016;2(1):120-144 [FREE Full text] [doi: [10.15761/crt.1000130](https://doi.org/10.15761/crt.1000130)]
72. Merchant RA, Tsoi CT, Tan WM, Lau W, Sandrasageran S, Arai H. Community-based peer-led intervention for healthy ageing and evaluation of the 'HAPPY' program. *J Nutr Health Aging* 2021;25(4):520-527 [FREE Full text] [doi: [10.1007/s12603-021-1606-6](https://doi.org/10.1007/s12603-021-1606-6)] [Medline: [33786571](https://pubmed.ncbi.nlm.nih.gov/33786571/)]
73. Gordon NP, Hornbrook MC. Older adults' readiness to engage with eHealth patient education and self-care resources: a cross-sectional survey. *BMC Health Serv Res* 2018 Mar 27;18(1):220 [FREE Full text] [doi: [10.1186/s12913-018-2986-0](https://doi.org/10.1186/s12913-018-2986-0)] [Medline: [29587721](https://pubmed.ncbi.nlm.nih.gov/29587721/)]
74. Gordon NP, Iribarren C. Health-related characteristics and preferred methods of receiving health education according to dominant language among Latinos aged 25 to 64 in a large Northern California health plan. *BMC Public Health* 2008 Sep 09;8:305 [FREE Full text] [doi: [10.1186/1471-2458-8-305](https://doi.org/10.1186/1471-2458-8-305)] [Medline: [18782454](https://pubmed.ncbi.nlm.nih.gov/18782454/)]
75. Bao GC, Dillon J, Jannat-Khah D, Besada M, Marianova A, Mathewos A, et al. Tai chi for enhanced inpatient mobilization: a feasibility study. *Complement Ther Med* 2019 Oct;46:109-115. [doi: [10.1016/j.ctim.2019.07.020](https://doi.org/10.1016/j.ctim.2019.07.020)] [Medline: [31519266](https://pubmed.ncbi.nlm.nih.gov/31519266/)]
76. Klein P, Picard G, Baumgarten J, Schneider R. Meditative movement, energetic, and physical analyses of three qigong exercises: unification of eastern and western mechanistic exercise theory. *Medicines (Basel)* 2017 Sep 23;4(4):69 [FREE Full text] [doi: [10.3390/medicines4040069](https://doi.org/10.3390/medicines4040069)] [Medline: [28946612](https://pubmed.ncbi.nlm.nih.gov/28946612/)]
77. Hung SH, Hwang CY, Chang CY. Is the Qi experience related to the flow experience? Practicing qigong in urban green spaces. *PLoS One* 2021 Jan 4;16(1):e0240180 [FREE Full text] [doi: [10.1371/journal.pone.0240180](https://doi.org/10.1371/journal.pone.0240180)] [Medline: [33395692](https://pubmed.ncbi.nlm.nih.gov/33395692/)]
78. Bauman A, Merom D, Bull FC, Buchner DM, Fiatarone Singh MA. Updating the evidence for physical activity: summative reviews of the epidemiological evidence, prevalence, and interventions to promote "active aging". *Gerontologist* 2016 Apr;56 Suppl 2:S268-S280. [doi: [10.1093/geront/gnw031](https://doi.org/10.1093/geront/gnw031)] [Medline: [26994266](https://pubmed.ncbi.nlm.nih.gov/26994266/)]
79. Kelley GA, Kelley KS. Meditative movement therapies and health-related quality-of-life in adults: a systematic review of meta-analyses. *PLoS One* 2015 Jun 8;10(6):e0129181 [FREE Full text] [doi: [10.1371/journal.pone.0129181](https://doi.org/10.1371/journal.pone.0129181)] [Medline: [26053053](https://pubmed.ncbi.nlm.nih.gov/26053053/)]
80. Klein PJ, Baumgarten J, Schneider R. Qigong and tai chi as therapeutic exercise: survey of systematic reviews and meta-analyses addressing physical health conditions. *Altern Ther Health Med* 2019 Sep;25(5):48-53. [Medline: [31221939](https://pubmed.ncbi.nlm.nih.gov/31221939/)]
81. Chen X, Li W. The effects of practicing "Health Qigong-WuQinxì" for 6 months on middle-aged and elderly people's ability of focusing attention. *J Beijing Sport Univ* 2006;29(10):1362-1363.
82. Shen M, Pi YL, Li Z, Song T, Jie K, Wang T, et al. The feasibility and positive effects of wuqinxì exercise on the cognitive and motor functions of patients with Parkinson's disease: a pilot study. *Evid Based Complement Alternat Med* 2021 Mar 29;2021:8833736 [FREE Full text] [doi: [10.1155/2021/8833736](https://doi.org/10.1155/2021/8833736)] [Medline: [33859712](https://pubmed.ncbi.nlm.nih.gov/33859712/)]
83. Buatois S, Miljkovic D, Manckoundia P, Gueguen R, Miget P, Vançon G, et al. Five times sit to stand test is a predictor of recurrent falls in healthy community-living subjects aged 65 and older. *J Am Geriatr Soc* 2008 Aug;56(8):1575-1577. [doi: [10.1111/j.1532-5415.2008.01777.x](https://doi.org/10.1111/j.1532-5415.2008.01777.x)] [Medline: [18808608](https://pubmed.ncbi.nlm.nih.gov/18808608/)]
84. Goldberg A, Chavis M, Watkins J, Wilson T. The five-times-sit-to-stand test: validity, reliability and detectable change in older females. *Aging Clin Exp Res* 2012 Aug;24(4):339-344. [doi: [10.1007/BF03325265](https://doi.org/10.1007/BF03325265)] [Medline: [23238309](https://pubmed.ncbi.nlm.nih.gov/23238309/)]
85. Miller SM, Taylor-Piliae RE. Effects of tai chi on cognitive function in community-dwelling older adults: a review. *Geriatr Nurs* 2014;35(1):9-19. [doi: [10.1016/j.gerinurse.2013.10.013](https://doi.org/10.1016/j.gerinurse.2013.10.013)] [Medline: [24252560](https://pubmed.ncbi.nlm.nih.gov/24252560/)]
86. Wayne PM, Walsh JN, Taylor-Piliae RE, Wells RE, Papp KV, Donovan NJ, et al. Effect of tai chi on cognitive performance in older adults: systematic review and meta-analysis. *J Am Geriatr Soc* 2014 Jan;62(1):25-39 [FREE Full text] [doi: [10.1111/jgs.12611](https://doi.org/10.1111/jgs.12611)] [Medline: [24383523](https://pubmed.ncbi.nlm.nih.gov/24383523/)]
87. Qin Y, Xia W, Huang W, Zhang J, Zhao Y, Fang M. The beneficial effect of traditional Chinese exercises on the management of obesity. *Evid Based Complement Alternat Med* 2020 Oct 1;2020:2321679 [FREE Full text] [doi: [10.1155/2020/2321679](https://doi.org/10.1155/2020/2321679)] [Medline: [33062003](https://pubmed.ncbi.nlm.nih.gov/33062003/)]
88. Siu PM, Yu AP, Chin EC, Yu DS, Hui SS, Woo J, et al. Effects of tai chi or conventional exercise on central obesity in middle-aged and older adults : a three-group randomized controlled trial. *Ann Intern Med* 2021 Aug;174(8):1050-1057. [doi: [10.7326/M20-7014](https://doi.org/10.7326/M20-7014)] [Medline: [34058100](https://pubmed.ncbi.nlm.nih.gov/34058100/)]
89. Physical activity guidelines for Americans. 2nd edition. U.S. Department of Health and Human Services. 2018. URL: [https://health.gov/paguidelines/second-edition/pdf/Physical\\_Activity\\_Guidelines\\_2nd\\_edition.pdf](https://health.gov/paguidelines/second-edition/pdf/Physical_Activity_Guidelines_2nd_edition.pdf) [accessed 2022-06-10]
90. Lavoie CJ, Zeidler MR, Martin JL. Sleep and aging. *Sleep Science Practice* 2018 Feb 23;2(1):3. [doi: [10.1186/s41606-018-0021-3](https://doi.org/10.1186/s41606-018-0021-3)]
91. Chang PS, Knobf T, Oh B, Funk M. Physical and psychological health outcomes of qigong exercise in older adults: a systematic review and meta-analysis. *Am J Chin Med* 2019;47(2):301-322. [doi: [10.1142/S0192415X19500149](https://doi.org/10.1142/S0192415X19500149)] [Medline: [30827152](https://pubmed.ncbi.nlm.nih.gov/30827152/)]
92. Lindquist R, Wyman JF, Talley KM, Findorff MJ, Gross CR. Design of control-group conditions in clinical trials of behavioral interventions. *J Nurs Scholarsh* 2007;39(3):214-221. [doi: [10.1111/j.1547-5069.2007.00171.x](https://doi.org/10.1111/j.1547-5069.2007.00171.x)] [Medline: [17760793](https://pubmed.ncbi.nlm.nih.gov/17760793/)]



93. Burnett-Zeigler I, Schuette S, Victorson D, Wisner KL. Mind-body approaches to treating mental health symptoms among disadvantaged populations: a comprehensive review. *J Altern Complement Med* 2016 Feb;22(2):115-124 [[FREE Full text](#)] [doi: [10.1089/acm.2015.0038](https://doi.org/10.1089/acm.2015.0038)] [Medline: [26540645](https://pubmed.ncbi.nlm.nih.gov/26540645/)]

## Abbreviations

**CHW:** community health worker

**MMI:** Meditative Movement Inventory

**PA:** physical activity

**SF-12:** 12-Item Short Form Health Survey

*Edited by A Mavragani; submitted 02.06.22; peer-reviewed by M Fotteler; comments to author 30.06.22; revised version received 28.07.22; accepted 22.08.22; published 13.09.22.*

*Please cite as:*

*Du Y, Patel N, Hernandez A, Zamudio-Samano M, Li S, Zhang T, Fernandez R, Choi BY, Land WM, Ullevig S, Estrada Coats V, Moussavou JMM, Parra-Medina D, Yin Z*

*Examining the Delivery of a Tailored Chinese Mind-Body Exercise to Low-Income Community-Dwelling Older Latino Individuals for Healthy Aging: Feasibility and Acceptability Study*

*JMIR Form Res* 2022;6(9):e40046

URL: <https://formative.jmir.org/2022/9/e40046>

doi: [10.2196/40046](https://doi.org/10.2196/40046)

PMID: [35997685](https://pubmed.ncbi.nlm.nih.gov/35997685/)

©Yan Du, Neela Patel, Arthur Hernandez, Maria Zamudio-Samano, Shiyu Li, Tianou Zhang, Roman Fernandez, Byeong Yeob Choi, William M Land, Sarah Ullevig, Vanessa Estrada Coats, Jessh Mondesir Mavoungou Moussavou, Deborah Parra-Medina, Zenong Yin. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 13.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Motive-Oriented, Personalized, Internet-Based Interventions for Depression: Nonclinical Experimental Study

Lara Bücker<sup>1</sup>, PhD; Thomas Berger<sup>2</sup>, PhD; Alina Bruhns<sup>1</sup>, MSc; Stefan Westermann<sup>3,4</sup>, PhD

<sup>1</sup>Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

<sup>2</sup>Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland

<sup>3</sup>Medical School Hamburg, University of Applied Sciences and Medical University, Hamburg, Germany

<sup>4</sup>Department of Psychology, University of California, Berkeley, CA, United States

**Corresponding Author:**

Lara Bücker, PhD

Department of Psychiatry and Psychotherapy

University Medical Center Hamburg-Eppendorf

Martinistrasse 52

Hamburg, 20246

Germany

Phone: 49 040 7410 55868

Email: [l.buecker@uke.de](mailto:l.buecker@uke.de)

## Abstract

**Background:** The low level of adherence in internet-based self-help interventions for depression suggests that in many existing programs, the motivational fit between the program and the user is unsatisfactory (eg, the user seeks autonomy, but the program provides directive guidance). Personalized, motive-oriented, self-help interventions could enable participants who interact with a program and its contents to have more engaging and less aversive experiences and thus increase adherence.

**Objective:** In an experimental study with a nonclinical analogue sample, we aimed to test the hypotheses that a better motivational person-program fit is linked with higher anticipated adherence, working alliance, and satisfaction with the program.

**Methods:** Motivational person-program fit was examined with respect to the 2 contrasting motives *being autonomous* and *being supported*. The hypotheses were tested by specifically varying the motivational person-program fit in a nonclinical sample (N=55), where participants were asked to work on, and subsequently evaluate, a limited set of individual pages of a self-help program with guidance (in the form of text messages) for depression. The sections of the self-help program were redesigned to either particularly address the autonomy motive or the support motive. For the quasi-experimental variation of the motivational person-program characteristics, we divided the 55 participants into 2 groups (autonomy group: n=27, 49%; support group: n=28, 51%) by screening method (using the Inventory of Approach and Avoidance Motivation), corresponding to the 2 motives. Both groups evaluated (in randomized order) 2 excerpts of the program—one that matched their motive (fit) and one that was contrary to it (no fit). Immediately after the evaluation of each excerpt, anticipated adherence, working alliance, and treatment satisfaction were assessed.

**Results:** Regarding *being supported*, the satisfaction with or violation of this motive had an impact on (optimal) anticipated adherence as well as working alliance and satisfaction with the intervention; a congruent person-program fit resulted in significantly higher anticipated adherence ( $t_{27}=3.00$ ;  $P=.006$ ), working alliance ( $t_{27}=3.20$ ;  $P=.003$ ), and satisfaction ( $t_{27}=2.86$ ;  $P=.008$ ) than a noncongruent fit. However, a similar impact could not be found for the motive *being autonomous*. Several correlations were found that supported our hypotheses (eg, for the congruent person-program fit autonomy motive and autonomy group, support satisfaction negatively correlated with optimal anticipated adherence).

**Conclusions:** This first experimental study gives reason to assume that motive orientation may have a positive influence on adherence, working alliance, and satisfaction in internet-based self-help interventions for depression and other mental disorders. Future studies should conduct randomized controlled trials with clinical samples and assess clinical outcomes.

(JMIR Form Res 2022;6(9):e37287) doi:[10.2196/37287](https://doi.org/10.2196/37287)

**KEYWORDS**

internet-based interventions; depression; adherence; motive orientation; personalization

## Introduction

### Background

To address the gap in the treatment of mental disorders [1], in the last decade, numerous internet-based interventions have been developed and investigated with regard to feasibility, acceptance, effectiveness, and side effects [2-4]. Internet-based interventions have some well-known advantages (eg, higher flexibility, resource saving, applicability, lower costs, and high levels of anonymity) over conventional treatment options such as psychotherapy and are therefore assumed to be able to address specific treatment barriers and narrow treatment gaps [5-7]. Internet-based interventions can be categorized as either self-guided (no additional individual therapeutic support), guided (with individual therapeutic support in the form of messages or contact through telephone), or blended (combination of internet-based programs and conventional psychotherapy) [8-10]. Most of these programs have been developed for anxiety disorders and depression [11], although programs for other mental disorders such as schizophrenia or pathological gambling are also addressed [12,13]. For programs with therapeutic guidance, meta-analyses have found effect sizes comparable to those for conventional face-to-face psychotherapy [2,7]. For self-guided programs, the effects reported are smaller; however, there are several advantages that speak for their use (eg, they provide increased access to treatment for those who do not meet the full criteria of a disorder, are cost-effective and resource saving, and there are no waiting times [14,15]).

### Adherence in Internet-Based Interventions

Although internet-based interventions have proven effective in reducing psychological symptoms, treatment adherence in such interventions is usually low, and treatment dropout rates are high [15,16]. Rates of dropout from internet-based interventions for depression can be as high as 75%, with a mean of 32% (SD 17%) [17]. Several predictors of treatment adherence in self-guided internet-based interventions for depression have been identified: male gender, low educational background, and comorbid anxiety disorders are associated with low adherence [18]. A more recent study found that a higher educational level, extraversion (personality trait), and participants' use of cognitive behavioral therapy skills predicted lower dropout risk, whereas technical difficulties and openness to experience predicted higher dropout risk [17]. Studies consistently found that only a minority of participants use self-guided interventions on a regular basis [13,16,19-21]. Findings further indicate that use is associated with effectiveness: those who regularly (ie, several times a week) work with the interventions do benefit best in terms of symptom reduction [22,23]. It is therefore crucial to investigate ways on how to enhance users' motivation to use such interventions frequently. Although Titov et al [23] found that sending reminders through automated emails facilitated treatment outcomes, a previous study on a smartphone app self-help intervention for depression aimed to increase use by sending daily smartphone reminders through push notifications but still found frequent use only for a few participants [22]. Consequently, it is necessary to look for new approaches that can sustainably increase adherence, possibly by personalizing interventions. Attempts to date to increase adherence are

invariably based on nontheoretical trials, and there is a lack of studies that test theoretically based approaches embedded in basic research.

### Personalization to Increase Adherence

There is widespread agreement that all forms of psychological interventions should be tailored to individual patient characteristics and needs [24]. We hypothesize that by adapting the content of internet-based interventions to specific patient characteristics as well as user needs and motives, adherence to these interventions can be increased. According to a recent meta-analysis, adapting psychotherapy to specific patient preferences is associated with fewer treatment dropouts and more positive treatment outcomes than providing a nonpreferred type or mode of treatment [25]. Another meta-analysis investigated the extent to which treatment outcomes are improved when therapists offer less directive treatments to patients with high reactance [26,27]. The results showed a large effect size (Cohen  $d=0.79$ ) and confirmed that individuals with high reactance had better outcomes when therapists adopted a reflective and nondirective attitude rather than a directive and authoritarian one. To a slightly lesser extent, a directive and authoritarian attitude for individuals with low reactance did yield better outcomes too.

When personalizing internet-based interventions, it is crucial to look beyond symptom severity [28]. Furthermore, personalization should not only mean that users receive personalized feedback on their exercises, but rather that the contents and elements of self-help programs are adapted, possibly based on a preliminary assessment. Berger et al [29] evaluated an internet-based, tailored, guided self-help treatment for social anxiety disorder, panic disorder with or without agoraphobia, and generalized anxiety disorder by comparing the tailored treatment with both a standardized disorder-specific internet-based treatment and a wait-list control group. The study found large effect sizes for both active groups as well as the control group (Cohen  $d=0.80$  and Cohen  $d=0.82$ , respectively), but no difference was found in effectiveness in all outcomes between the 2 active groups. Furthermore, a Swedish research group developed internet-based interventions for anxiety disorders and depression that tailor the content to the individual symptoms of the user and evaluated the approach in several trials [30-32]. The group found that the tailored approach was superior to the approaches used for the active control groups (web-based discussion group [30] and standardized, nontailored internet-based treatment [32]). A more recent study compared a tailored internet-based intervention for depression and anxiety in older adults with weekly general support and found a higher reduction of anxiety (Cohen  $d=0.50$ ) in the tailored approach [33]. Moritz et al [34] evaluated a web-based self-help program for obsessive-compulsive disorder and examined whether a version tailored to individual problems would produce greater effects than the full program, but they did not find better outcomes for the tailored approach. In most of these programs, tailoring consisted of selecting and sequencing existing modules specifically for the user based on information obtained from a web-based assessment. Another option could be to adapt internet-based interventions to individual motives of users [35].

## Motive-Oriented Internet-Based Interventions

According to the *consistency theory*, human functioning can be understood as need driven, and approach and avoidance motives are assumed to promote satisfaction of the basic needs for control and orientation, pleasure, attachment, and self-enhancement, as well as to prevent frustration or violation of these needs [36-38]. In this context, *incongruence* can be understood as insufficient motive satisfaction in interaction with the environment [39]. In psychotherapy, therapists derive the most important motives of their patients from their behavior (*bottom-up*) with the help of plan analysis and proactively take them into account when building and maintaining the therapy relationship [40].

From a motivational perspective, internet-based interventions can be understood as environments that individuals access to satisfy their individual needs or to prevent the needs from being violated [35,36]. The low level of adherence in self-guided internet-based interventions for depression suggests that in many existing programs the motivational fit between the program and the user is unsatisfactory (eg, the user seeks autonomy, although the program provides much support and guidance in a directive style). Personalized, motive-oriented, self-guided interventions could enable individuals who interact with the program and its contents to have more engaging and less aversive experiences and thus increase treatment adherence [41,42].

## Objectives of This Study

Within a nonclinical experimental study, excerpts of a web-based intervention for depression were adapted to the following 2 potentially opposing motives: *being autonomous* and *being supported*. Using these 2 motives, we tested the hypotheses that a better motivational person-program fit is associated with (1) higher anticipated adherence, (2) working alliance, and (3) satisfaction with the program.

## Methods

### Study Design

This study was based on an experimental design. The independent variables were (1) group (groups screened with the Inventory of Approach and Avoidance Motivation based on terciles: autonomy group [AutGrp]=autonomy high, >2/3, and support low, <1/3, vs support group [SuppGrp]=support high, >2/3, and autonomy low, <1/3) and (2) motive orientation through 2 versions of excerpts of a self-help internet-based intervention (autonomy condition [AutCond] and support condition [SuppCond]) tailored to the 2 groups. A motivational fit existed in the conditions AutCond-AutGrp and SuppCond-SuppGrp, whereas a discrepancy existed in the conditions SuppCond-AutGrp and AutCond-SuppGrp. The dependent variables were motivational incongruence during program use, expected adherence, working alliance, and treatment satisfaction.

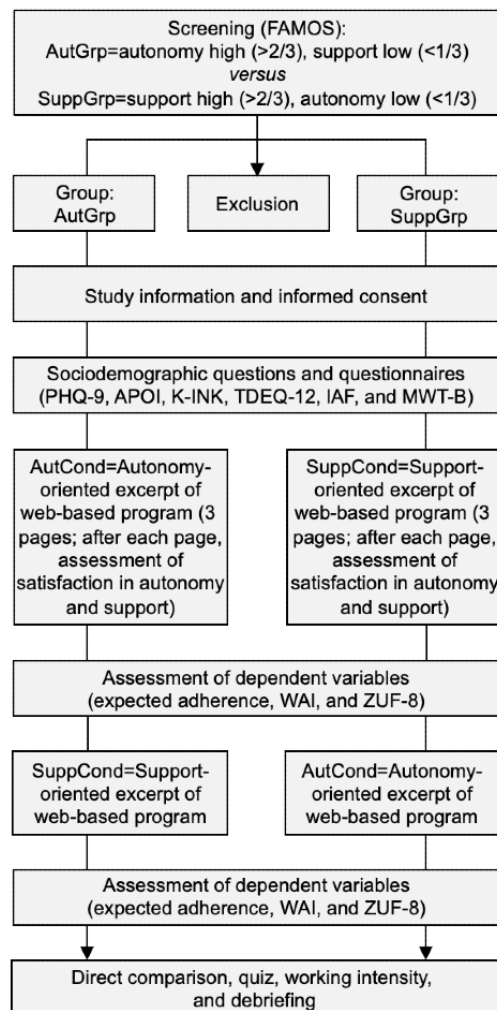
## Procedures

The study was performed at the University Medical Center Hamburg-Eppendorf (Germany), and the assessment was conducted on the web using the survey software Unipark (Questback). Refer to [Figure 1](#) for a detailed procedure of the assessment.

On the first page of the excerpt of the self-help intervention, prospective participants received basic information about the study (participants were informed that they would see different versions of excerpts of an internet-based intervention that they would have to evaluate; no information on the motive adaptation was provided before the study participation) and were instructed that they would be briefly screened to determine whether they were suitable to participate in the study. In addition, to prevent fake participation by computer programs, a simple arithmetic problem had to be solved to get to the second page. Prospective participants were then screened using the Questionnaire for the Analysis of Motivation Schemas (Fragebogen zur Analyse Motivationaler Schemata [FAMOS]) [43] to form the 2 groups (high values for autonomy and low values for support vs high values for support and low values for autonomy). Candidates who did not achieve such suitable scores (refer to the *Study Design* section) on this questionnaire were excluded from further participation. The study participation then lasted approximately 1 hour, and all study participants were rewarded with a €10 (US \$10.30) Amazon voucher for their participation (given only to those who passed the screening and completed the entire survey). At the end of the study, participants were informed in detail about the background and the course of the study by means of a printable information page. Subsequently, an electronic declaration of informed consent for study participation and data collection was requested.

This was followed by the administering of the questionnaires (refer to the *Measures* section) as well as the processing of both versions of the 3-page excerpts of the self-help program (refer to the *Excerpts of the Self-help Internet-Based Intervention for Depression* section) and their evaluation. After each page, the participants were asked for their motivational incongruence, mood, and arousal as well as 2 questions on how they would rate the quality and comprehensibility of the page. The order of the excerpt versions was randomized (AutCond-SuppCond or SuppCond-AutCond). After the evaluation of each excerpt, the dependent variables (expected adherence, working alliance, and treatment satisfaction) were assessed. At the end of the survey, 4 knowledge questions about the content of the self-help excerpts were asked, and the participants received a short debriefing on the rationale of the study. Subsequently, participants could enter their email address for receiving the voucher. As the email address was processed and stored independently of all other information, the study data were completely anonymous at all times. Accordingly, participants were informed that they could not subsequently request the deletion of their data. Participants were notified that they were not participating in an intervention study.

**Figure 1.** Flowchart of the assessment procedure. APOI: Attitudes Toward Psychological Online Interventions; AutCond: autonomy condition; AutGrp: autonomy group; FAMOS: Fragebogen zur Analyse Motivationaler Schemata (Inventory of Approach and Avoidance Motivation); IAF: Index of Autonomous Functioning; K-INK: Kurzversion des Inkongruenzfragebogen (Incongruence Questionnaire, short version); MWT-B: Multiple-Choice Vocabulary Intelligence Test; PHQ-9: Patient Health Questionnaire-9, depression module; SuppCond: support condition; SuppGrp: support group; TDEQ-12: Theoretical Depressive Experiences Questionnaire, short version; WAI: Working Alliance Inventory; ZUF-8: Zufriedenheitsfragebogen (Client Satisfaction Questionnaire).



## Sample Size

The power analysis for calculating the sample size for an ANOVA was conducted using G\*Power (Heinrich Heine University Düsseldorf) [44], and it revealed a sample size of  $n=54$  to detect a medium effect of Cohen  $f=0.25$ , with Cronbach  $\alpha=.05$  and a power of 0.95.

## Recruitment

The sample was recruited through different student, psychology, and study announcement internet forums, and Facebook groups. The recruitment took place from May 16, 2019, to October 19, 2019.

## Eligibility Criteria

Those interested could participate if German was their native language, and they were aged  $\geq 18$  years. A current self-reported mental illness (eg, depression) was a criterion for exclusion. Filters were used to automatically prevent unsuitable prospective participants from participating (eg, age  $< 18$  years or self-reported mental illness).

## Ethics Approval

The local psychological ethics committee of the Center for Psychosocial Medicine of the University Medical Center Hamburg-Eppendorf approved the project (LPEK-0033; April 28, 2019). All participants provided informed consent on the web before participating in the study. The study was conducted in accordance with the Declaration of Helsinki, and the authors assert that all procedures contributing to this work comply with the ethics standards of relevant national and institutional guidelines.

## Excerpts of the Self-help Internet-Based Intervention for Depression

The excerpts that have been adapted to address the 2 motives (autonomy and support) were derived from an existing self-guided internet-based intervention with guidance in the form of text messages for depression called *MOOD* [20]. For both motives, 3 pages of the intervention were adapted to satisfy the need for autonomy (AutCond) and the need for support (SuppCond). The pages included an introduction to the program, psychoeducation on cognitive strategies and automatic thoughts,

as well as information on ABC schemas. In both conditions, therapeutic support was provided in the form of text messages: the nature of the support differed in that in the SuppCond, the guide was proactive and made contact with the user, whereas

in the AutGrp, the guide was available upon request, that is, the initiative came from the users. Examples of how the contents were adapted are presented in [Table 1](#).

**Table 1.** Examples of motive-oriented adaptations.

	Autonomy condition	Support condition
Interventions as means for motive satisfaction	“The ability to become aware of automatic thoughts and thus to create more freedom of thought can be acquired in this unit.” (Module: ABC schema)	“In this way, automatic thoughts can separate you from the feeling of being connected to others, for example. The ability to become aware of automatic thoughts such as ‘He doesn’t like me anymore’ and to question them specifically is taught in this unit.”
Motive-oriented formulations	“Of course you are not told what to think. On the contrary, you remain responsible for questioning and developing your own thinking.”	“Regardless of whether it is easy for you to learn and practice the techniques in this unit or not, your personal guide will contact you regularly and ask how you are doing.”
Motive-oriented functions	Privacy settings (eg, whether guide has access to worksheets)	Accessibility guide (eg, text fields for direct questions to the guide)
Motive-oriented guidance	“If I do not receive a message from you, I assume that you will be able to manage processing of the contents of this self-help program independently.”	“Even if you do not contact me, I will check at least once a week how far you have progressed in the program and what you have entered in the worksheets to give you personal feedback with suggestions.”

## Measures

### Overview

Data were collected on the web through the survey software Unipark. The following data were collected: motives (basis for screening and group formation; German version of the Inventory of Approach and Avoidance Motivation: FAMOS), motivational incongruence (Incongruence Questionnaire, short version [Kurzversion des Inkongruenzfragebogen (K-INK)]), attitudes regarding web-based interventions (Attitudes Toward Psychological Online Interventions [APOI]), dependency (Theoretical Depressive Experiences Questionnaire, short version [TDEQ-12]), autonomy (Index of Autonomous Functioning [IAF]), depressiveness (Patient Health Questionnaire-9, depression module [PHQ-9]), the level of intelligence (Multiple-Choice Vocabulary Intelligence Test [MWT-B]), expected adherence, working alliance, and satisfaction with the program (Client Satisfaction Questionnaire [Zufriedenheitsfragebogen (ZUF-8)]).

### Primary Outcome Measure

To measure the anticipated adherence, the following questions were asked about expected time investment and frequency of log-ins:

1. *How much time would you like to actively spend in the full version of the program (in hours and minutes; optimal anticipated adherence)?*
2. *How much time would you realistically actually spend in the full version of the program (in hours and minutes; realistic anticipated adherence)?*
3. *In how many of the 10 weeks would you log in to the program?*
4. *In the weeks with log-in: On average, how often would you log in to the program?*

5. *If you would still have the opportunity to log in after the 10 weeks in the program: How often would you do this in a year?*

Participants provided the answers in free-text fields, where only whole numbers could be entered.

### Secondary Outcome Measures

#### Working Alliance Inventory

The Working Alliance Inventory (WAI) [45] is a self-rating instrument for measuring the quality of alliance based on the pantheoretical, tripartite conceptualization of the therapeutic alliance (agreement on treatment goals, agreement on the tasks of the therapy, and development of a therapeutic bond). In this study, the short version of WAI [46,47], which consists of 12 items, was used. These 12 items were adapted to correspond to an internet-based intervention. The reference was no longer to a therapist but to a guide or contact person whose availability was announced on the excerpts of the web-based program (but with whom there was no actual interaction). For each of the three subscales (goals, tasks, and bond), 4 items were rated on a 5-point Likert scale ranging from 1=*rarely* to 5=*always*. The internal consistency of the subscales is good (Cronbach  $\alpha$ =.81-.91), and it is excellent for the total scale (Cronbach  $\alpha$ =.90-.93 [48]).

#### Patient Satisfaction (Measured Using the ZUF-8)

The ZUF-8 [49] is the German version of the Client Satisfaction Questionnaire. This self-assessment questionnaire consists of 8 items that serve to assess patients’ treatment satisfaction (eg, psychotherapy). The 8 items can be rated on a 4-point rating scale (*excellent, good, lessgood, and bad*). A total score ranging from 8 to 32 can be achieved, with high scores indicating a high degree of satisfaction. The internal consistency ranges from Cronbach  $\alpha$ =.87 to Cronbach  $\alpha$ =.93 [50].

### ***Instruments Used for the Formation and Verification of the Two Groups***

#### **FAMOS: Inventory of Approach and Avoidance Motivation**

The self-report questionnaire FAMOS [43] captures the motives of patients undergoing psychotherapy in terms of central components of motivational schemes and was used for the formation of the 2 groups. The motives are assessed as approach motives (14 scales, eg, intimacy and attachment, status, and performance) and avoidance motives (9 scales, eg, loneliness and separation, disregard, and failure) with a total of 94 items. The items are rated on 5-point Likert scales from *not important at all* to *extremely important* or *not bad at all* to *extremely bad*. In this study, only the approach (9 items) and avoidance (10 items) scales for the 2 motives *autonomy* and *support* were assessed. The internal consistency of the individual scales varies between Cronbach  $\alpha=.37$  and Cronbach  $\alpha=.93$  [43].

#### **TDEQ-12: Used to Measure Depressive Experiences**

The TDEQ-12 [51], one of the short versions of the questionnaire on depressive experiences [52], is a self-reporting questionnaire that intends to distinguish between dependence and self-criticism. The questionnaire consists of 12 items that are scored on a 7-point Likert scale ranging from 1=*strongly disagree* to 7=*strongly agree*. The questionnaire has satisfactory psychometric properties. The dependency factor was made up of 5 items that had internal consistency with a Cronbach  $\alpha=.77$ , whereas the internal consistency of the self-criticism factor consisting of 7 items had a Cronbach  $\alpha=.78$  [53]. In our study, only the 5 items assessing dependency were used to check whether the 2 groups differed significantly in dependency.

#### **IAF: Used to Measure Trait Autonomy**

The IAF [54] is a self-rating questionnaire to measure trait autonomy based on three theoretically derived subscales assessing authorship or self-congruence, interest-taking, and low susceptibility to control. The questionnaire consists of 5 items per scale that can be answered on a 5-point rating scale (1=*not at all true*, 2=*a bit true*, 3=*somewhat true*, 4=*mostly true*, and 5=*completely true*). The internal consistency of the scale is good (Cronbach  $\alpha=.82$  [54]). The IAF was used to check whether the 2 groups differed significantly in autonomy.

### ***Instruments Used for Testing Group Comparability***

#### **APOI: Used to Measure Attitudes Toward Psychological Web-Based Interventions**

The APOI [55] is a self-assessment questionnaire for assessing attitudes toward psychological web-based interventions covering four dimensions: (1) skepticism and risk perception, (2) trust in therapeutic efficacy, (3) perception of deficits in mechanization, and (4) perception of advantages of anonymity. It consists of 16 items that can be rated on a 5-point rating scale ranging from 1=*do not agree at all* to 5=*fully agree*. A higher total score indicates a more positive attitude toward psychological web-based interventions. All 4 dimensions are equally weighted. The APOI's internal consistency has a Cronbach  $\alpha=.77$  [55].

#### **K-INK: Used to Measure Incongruence**

Incongruence (unsatisfactory fit of motivational goal and actual experience) was assessed with the short version of the INK [56], a self-rating questionnaire with originally 94 items on 14 approach scales and 9 avoidance scales. The INK has satisfactory reliability (Cronbach  $\alpha=.72-.87$ ) and validity. Item contents and scale structure are derived from the FAMOS. Although the FAMOS measures the intensity of motives (importance or being bad), the INK measures the degree of insufficient implementation of these motives (satisfaction with the implementation of approach motives or the occurrence of avoidance motives). The short version (K-INK) contains only the 23 items with the highest item total correlations. Answers can be given on a 5-point Likert scale ranging from 1=*much too little* to 5=*completely sufficient*.

#### **PHQ-9: Used to Measure Depressive Symptoms**

Depressive symptoms were assessed with the PHQ-9 [57], which is a self-rating questionnaire that consists of 9 items on depressive symptoms over the past 2 weeks. The items can be answered on a 4-point Likert scale ranging from 0=*not at all* to 3=*nearly every day*. The questionnaire can assist the diagnosis of major depression according to Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, criteria. Its internal consistency ranges from Cronbach  $\alpha=.86$  to Cronbach  $\alpha=.89$  [58].

#### **MWT-B: Used to Measure General Intelligence Levels**

The MWT [59] is a performance test to measure general intelligence levels, specifically crystallized intelligence. The test includes 2 versions—A and B. Both versions consist of a total of 37 lines with 5 terms each. In each line only 1 of these 5 terms is a real word; this is the one to be found out and marked by the participant. The other terms are fictitious constructions. Test results correlate quite well with the global IQ of healthy adults (median of  $r=0.72$  [59]). In our study, we randomly presented each participant with only 1 item (a line of 5 terms) to exclude participants whose intelligence level was too low per the exclusion criteria.

### **Statistical Analyses**

Group differences in sociodemographic characteristics and assessed questionnaires were determined using Welch 2-tailed *t* tests and chi-square tests. The Shapiro-Wilk test was used to test for normality of data. Logarithmic (log) transformation was applied for skewed data. The main hypothesis was tested using an ANOVA with repeated measurements (within-subject: motive orientation and between-subject: motive). Multiple comparisons were adjusted with Bonferroni correction. Within-group differences were analyzed with paired sample *t* tests and between-group differences with independent sample *t* tests. Correlative relationships were analyzed using Pearson correlation coefficients.

## **Results**

### **Sample Characteristics**

In total, 55 participants were included in the analyses. No participant was excluded from the study on account of

intelligence levels (based on the MWT-B). Sample characteristics are depicted in Table 2. The participants' mean age was 27.27 (8.18) years. The majority of the participants were women (47/55, 85%), and almost half were single (27/55, 49%). The 2 groups (AutGrp: 27/55, 49%, and SuppGrp: 28/55, 51%) did not differ in sociodemographic characteristics, depressive episodes in the past (lifetime), current depressive symptoms (PHQ-9), attitudes toward web-based interventions (APOI), or motivational incongruence (K-INK).

The formation of 2 groups based on the screening with the FAMOS questionnaire worked very well. As intended, the 2 groups differed significantly in their responses to independent questionnaires of autonomy (IAF;  $P=.01$ ; Cohen  $d=-0.786$ , 95% CI  $-1.335$  to  $-0.237$ ) and dependence (TDEQ-12;  $P<.001$ ; Cohen  $d=1.321$ , 95% CI  $0.738-1.905$ ) with large effect sizes. Thus, the AutGrp had high levels of autonomy and low levels of dependence compared with the SuppGrp, and the check of the group formation can be deemed successful.

**Table 2.** Sociodemographic characteristics of the sample (N=55).

	Autonomy group (n=27)	Support group (n=28)	<i>t</i> test ( <i>df</i> )	Chi-square ( <i>df</i> )	<i>P</i> value
Age (years), mean (SD)	28.48 (9.44)	26.11 (6.71)	1.07 (53)	N/A <sup>a</sup>	.29
Sex (female), n (%)	22 (81)	25 (89)	N/A	0.67 (2)	.41
Marital status (single), n (%)	15 (56)	12 (43)	N/A	0.89 (3)	.35
School education (years), mean (SD)	12.11 (0.80)	12.29 (1.08)	-0.68 (53)	N/A	.50
Past depression, n (%)	4 (15)	6 (21)	N/A	0.40 (1)	.53
Autonomy (IAF <sup>b</sup> ), mean (SD)	4.13 (0.59)	3.65 (0.63)	2.94 (53)	N/A	.01
Dependence (TDEQ-12 <sup>c</sup> ), mean (SD)	2.67 (1.23)	4.59 (1.64)	-4.92 (53)	N/A	<.001
Depressive syndrome (PHQ-9 <sup>d</sup> ), mean (SD)	13.54 (3.87)	15.03 (4.41)	-1.33 (53)	N/A	.19
Attitudes toward web-based interventions (APOI <sup>e</sup> ), mean (SD)	52.26 (8.14)	49.29 (7.67)	1.39 (53)	N/A	.17
Motivational incongruence (K-INK <sup>f</sup> ), mean (SD)	2.01 (0.46)	2.27 (0.66)	-1.72 (53)	N/A	.09

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

<sup>b</sup>IAF: Index of Autonomous Functioning.

<sup>c</sup>TDEQ-12: Theoretical Depressive Experiences Questionnaire, short version.

<sup>d</sup>PHQ-9: Patient Health Questionnaire-9, depression module.

<sup>e</sup>APOI: Attitudes Toward Psychological Online Interventions.

<sup>f</sup>K-INK: Kurzversion des Inkongruenzfragebogen (Incongruence Questionnaire, short version).

## Manipulation Checks

The violation and satisfaction of the 2 motives *being autonomous* and *being supported* for the different fits of group and condition are presented in Figure 2. Focusing on the incongruent and congruent fits (ie, mismatch and match) for the AutGrp in autonomy violation, the incongruent fit SuppCond-AutGrp reported a higher autonomy violation (mean 1.73, SD 0.72) than the congruent fit AutCond-AutGrp (mean 1.44, SD 0.65;  $t_{27}=2.36$ ;  $P=.03$ ), which is in line with our expectations. The results of the interaction effect of the repeated measures ANOVA showed that all 4 combinations did not differ significantly in autonomy violation ( $F_{1,53}=0.06$ ;  $P=.80$ ). However, the main effects for condition ( $F_{1,53}=4.95$ ;  $P=.03$ ) and group ( $F_{1,53}=5.90$ ;  $P=.02$ ) were significant. Overall, the SuppCond (mean 1.97, SE 0.12) and the SuppGrp (mean 2.09, SE 0.15) experienced higher levels of autonomy violation than

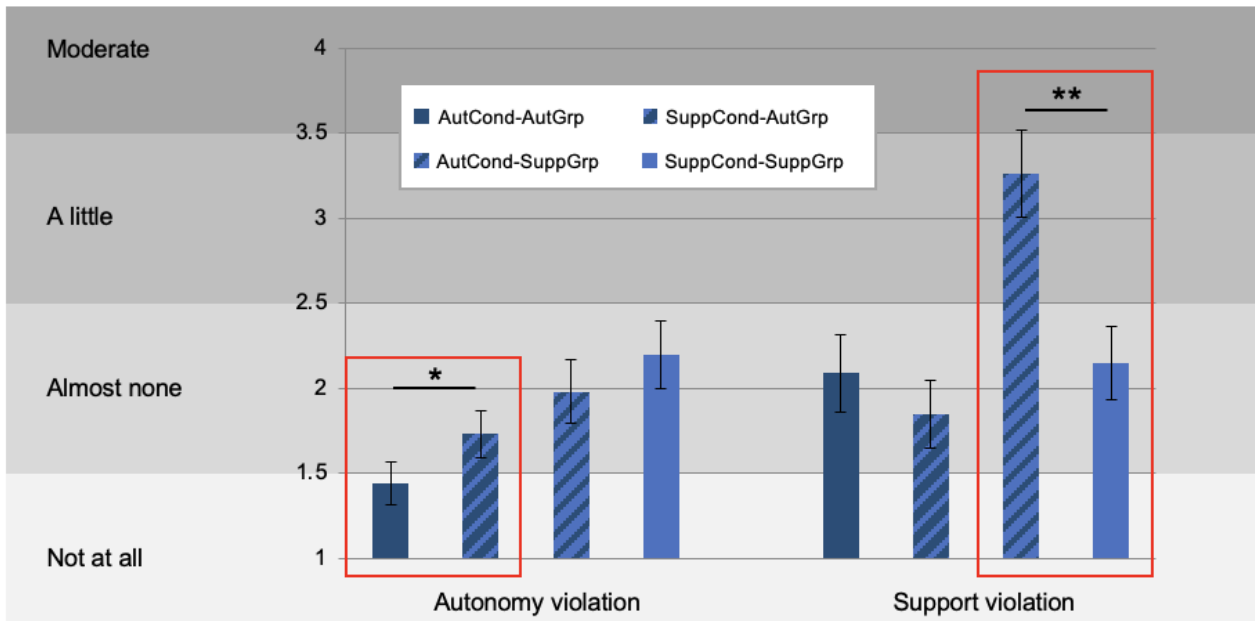
the AutCond (mean 1.71, SE 0.12) and the AutGrp (mean 1.57, SE 0.15).

For support violation, AutCond-SuppGrp reported a higher support violation (mean 3.26, SD 1.36) than the congruent fit SuppCond-SuppGrp (mean 2.15, SD 1.14;  $t_{28}=4.50$ ;  $P<.001$ ), which is also consistent with our expectations. For support violation, the interaction effect was significant ( $F_{1,53}=7.37$ ;  $P=.009$ ), as were both the main effects for condition ( $F_{1,53}=17.44$ ;  $P<.001$ ) and group ( $F_{1,53}=7.12$ ;  $P=.01$ ). For condition, support violation was higher in the AutCond (mean 2.67, SE 0.17) than in the SuppCond (mean 2.00, SE 0.15). However, for group, support violation was higher in the SuppGrp (mean 1.97, SE 0.20) than in the AutGrp (mean 2.71, SE 0.19). The interaction term stems from a significant difference among the conditions in the SuppGrp ( $t_{27}=5.00$ ;  $P<.001$ ) and a nonsignificant difference among the conditions in the AutGrp ( $t_{26}=1.01$ ;  $P=.32$ ).

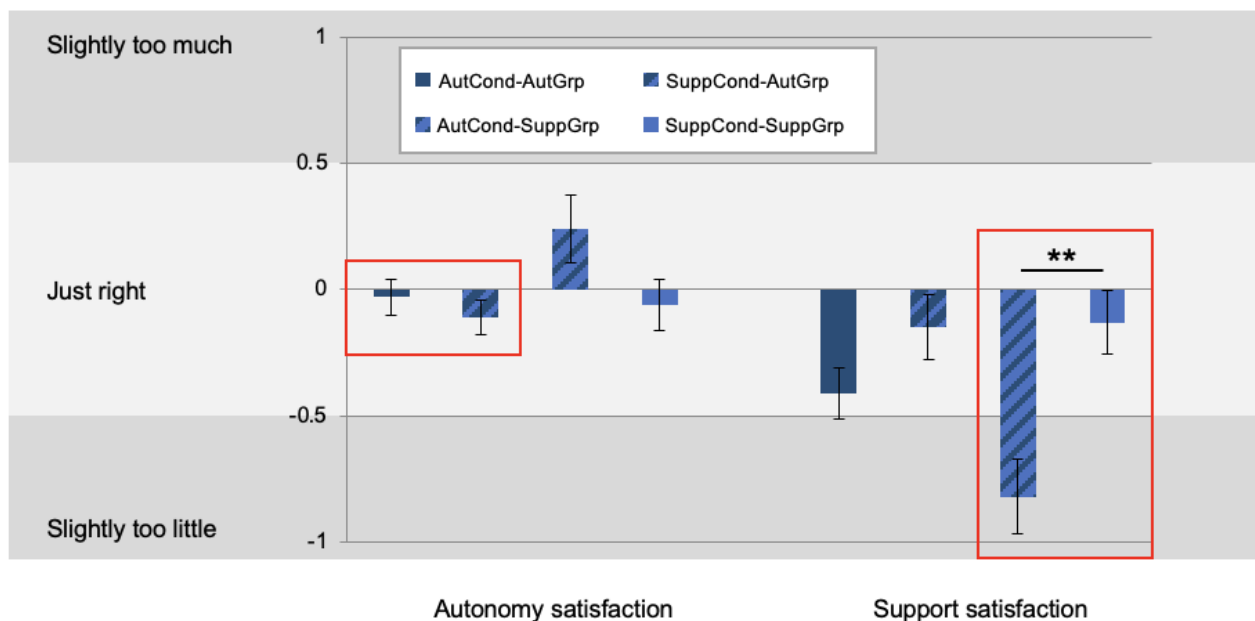


**Figure 2.** Manipulation checks. (A) Autonomy and support violation per group and condition. (B) Autonomy and support satisfaction per group and condition. AutCond: autonomy condition; AutGrp: autonomy group; SuppCond: support condition; SuppGrp: support group. \* $P < .05$ , \*\* $P < .001$ .

**A**



**B**



Regarding autonomy satisfaction, as expected, autonomy in the incongruent fit SuppCond-AutGrp was *slightly too little* (mean  $-0.11$ , SD  $0.36$ ), and the congruent fit AutCond-AutGrp reported an autonomy satisfaction that was almost *just right* (mean  $-0.03$ , SD  $0.37$ ). However, this difference was not significant ( $t_{27}=0.83$ ;  $P=.41$ ). The interaction effect ( $F_{1,53}=2.21$ ;  $P=.14$ ) and the main effect for group ( $F_{1,53}=1.78$ ;  $P=.19$ ) for autonomy satisfaction were not significant, but the main effect for condition was significant ( $F_{1,53}=4.95$ ;  $P=.03$ ). Specifically, satisfaction

regarding autonomy was closer to *just right* in the SuppCond (mean  $-0.09$ , SE  $0.06$ ) than in the AutCond (mean  $0.10$ , SE  $0.08$ ).

Finally, for support satisfaction, the incongruent fit AutCond-SuppGrp (mean  $-0.82$ , SD  $0.79$ ) had the lowest ratings (*slightly too little* support). In the congruent fit SuppCond-SuppGrp, the rating was the closest to *just right* (mean  $-0.06$ , SD  $0.54$ ). This difference was significant ( $t_{28}=4.34$ ;

$P < .001$ ). The interaction effect here was significant ( $F_{1,53}=4.88$ ;  $P=.03$ ), as was the main effect for condition ( $F_{1,53}=3.74$ ;  $P < .001$ ); however, the effect for group was not significant ( $F_{1,53}=1.97$ ;  $P=.17$ ). Overall, the satisfaction regarding support was closer to *just right* in the SuppCond (mean  $-0.14$ , SE  $0.08$ ) than in the AutCond (mean  $-0.61$ , SE  $0.09$ ). The significant interaction term stems from a significant difference between the groups in the AutCond ( $t_{53}=2.27$ ;  $P=.03$ ) and a nonsignificant difference between the groups in the SuppCond ( $t_{53}=0.10$ ;  $P=.92$ ).

### Anticipated Adherence

The results of the anticipated adherence (in anticipated minutes spent with the program; optimal and realistic), working alliance, and satisfaction are shown in Table 3 and Figure 3. The highest anticipated adherence (both optimal and realistic) was found for the combinations SuppCond-SuppGrp (optimal adherence: mean  $319.39$ , SD  $333.60$ , and realistic adherence: mean  $201.82$ , SD  $230.96$ ; refer to Table 3 for all log-transformed data) and AutCond-SuppGrp (optimal adherence: mean  $280.29$ , SD  $356.16$ , and realistic adherence: mean  $209.32$ , SD  $305.01$ ); the lowest anticipated adherence was found for the combination AutCond-AutGrp (optimal adherence: mean  $193.59$ , SD  $180.82$ ; realistic adherence: mean  $121.15$ , SD  $97.93$ ). The results of the interaction effect of the repeated measures ANOVA showed that the combinations did not significantly differ for anticipated adherence (log-optimal adherence:  $F_{1,53}=0.88$ ;  $P=.35$ , and log-realistic adherence:  $F_{1,53}=0.59$ ;  $P=.45$ ). In addition, no main

effects for group were found (log-optimal adherence:  $F_{1,53}=0.32$ ;  $P=.86$ , and log-realistic adherence:  $F_{1,53}=0.001$ ;  $P=.98$ ). For condition, a main effect was present for optimal anticipated adherence ( $F_{1,53}=6.49$ ;  $P=.01$ ), with higher ratings in the SuppCond (mean  $5.07$ , SE  $0.14$ ) than in the AutCond (mean  $4.89$ , SE  $0.15$ ), but a main effect was not present for realistic anticipated adherence ( $F_{1,53}=2.04$ ;  $P=.16$ ).

The results of the paired sample  $t$  tests indicated significant within-group differences for optimal anticipated adherence in the SuppGrp ( $t_{27}=3.00$ ;  $P=.006$ ). No significant within-group differences could be found for realistic anticipated adherence in the AutGrp.

Several correlative relationships between motive satisfaction or dissatisfaction and anticipated (optimal and realistic) adherence could be found. For the congruent person-program fit AutCond-AutGrp, support satisfaction negatively correlated with optimal anticipated adherence ( $r=-0.40$ ;  $P=.04$ ) and, in a matching result, support dissatisfaction positively correlated with optimal anticipated adherence ( $r=0.43$ ;  $P=.02$ ). In addition, for the incongruent fit SuppCond-AutGrp, support satisfaction negatively correlated with optimal anticipated adherence ( $r=-0.59$ ;  $P=.001$ ), and in another matching result, support dissatisfaction negatively correlated with optimal anticipated adherence ( $r=0.58$ ;  $P=.002$ ). For realistic adherence, in the congruent person-program fit AutCond-AutGrp, autonomy satisfaction correlated with realistic anticipated adherence ( $r=0.46$ ;  $P=.02$ ).

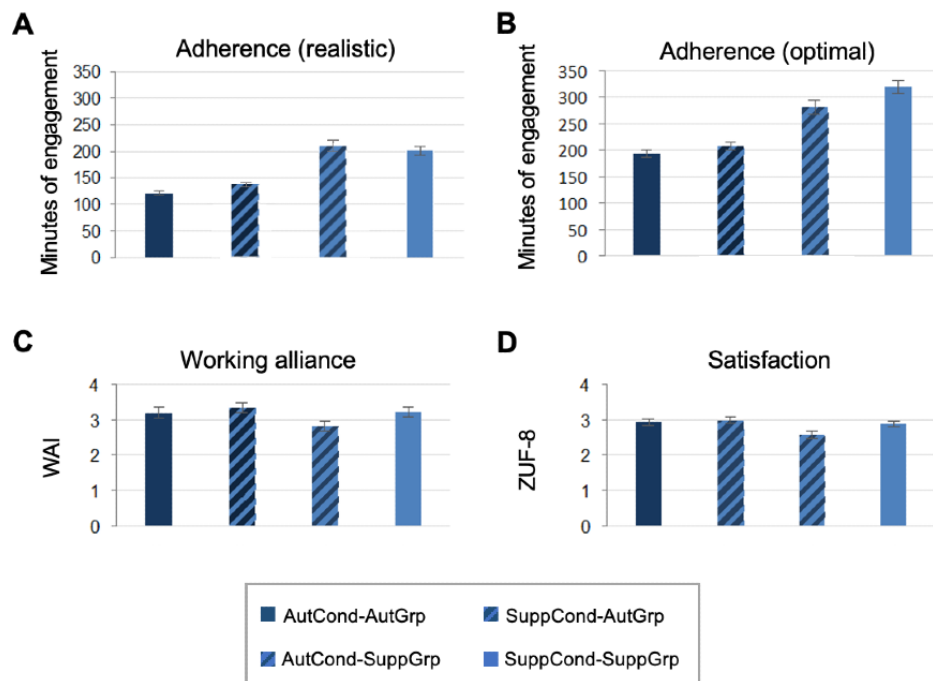
**Table 3.** Anticipated adherence (optimal and realistic) shown as expected time (in minutes) spent with the program; working alliance was measured with the Working Alliance Inventory (WAI), and satisfaction with the program was measured with the Client Satisfaction Questionnaire (Zufriedenheitsfragebogen [ZUF-8];  $N=55$ ).

Condition and variables	Autonomy group (n=27)		Support group (n=28)	
	Values, mean (SD)	Log-transformed mean (SD)	Values, mean (SD)	Log-transformed mean (SD)
<b>Autonomy condition</b>				
Adherence (optimal)	<i>193.59 (180.82)</i> <sup>a</sup>	<i>4.89 (0.90)</i>	280.29 (356.16)	4.88 (1.27)
Adherence (realistic)	<i>121.15 (97.93)</i>	<i>4.52 (0.84)</i>	209.32 (305.01)	4.43 (1.35)
WAI	<i>3.19 (0.80)</i>	N/A <sup>b</sup>	2.82 (0.73)	N/A
ZUF-8	<i>2.93 (0.53)</i>	N/A	2.56 (0.59)	N/A
<b>Support condition</b>				
Adherence (optimal)	207.85 (182.45)	5.01 (0.81)	<i>319.39 (333.60)</i>	<i>5.13 (1.25)</i>
Adherence (realistic)	138.00 (128.95)	4.59 (0.83)	<i>201.82 (230.96)</i>	<i>4.67 (1.20)</i>
WAI	3.34 (0.79)	N/A	<i>3.21 (0.66)</i>	N/A
ZUF-8	2.97 (0.49)	N/A	<i>2.88 (0.45)</i>	N/A

<sup>a</sup>Italics indicate congruence (normal font indicates incongruence).

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

**Figure 3.** (A) Anticipated realistic adherence per group and condition. (B) Anticipated optimal adherence per group and condition. (C) Working alliance per group and condition. (D) Satisfaction per group and condition. AutCond: autonomy condition; AutGrp: autonomy group; SuppCond: support condition; SuppGrp: support group; WAI: Working Alliance Inventory; ZUF-8: Zufriedenheitsfragebogen (Client Satisfaction Questionnaire).



### Working Alliance and Satisfaction

For working alliance (with the program and a guide whose availability was announced), neither a significant interaction effect ( $F_{1,53}=2.12$ ;  $P=.15$ ) nor a significant main effect for group ( $F_{1,53}=1.84$ ;  $P=.18$ ) was shown, but a significant main effect for condition was present ( $F_{1,53}=11.23$ ;  $P=.001$ ). Working alliance was descriptively highest for the combination SuppCond-AutGrp (mean 3.34, SD 0.79) and lowest for the combination AutCond-SuppGrp (mean 2.82, SD 0.73). Likewise, no significant interaction effect was found for satisfaction ( $F_{1,53}=3.48$ ;  $P=.07$ ). Here, the main effect for group bordered on significance ( $F_{1,53}=3.77$ ;  $P=.06$ ), whereas the main effect for condition was significant ( $F_{1,53}=5.56$ ;  $P=.02$ ). Satisfaction was significantly higher in the SuppCond (mean 2.92, SE 0.06) than in the AutCond (mean 2.75, SE 0.08). Satisfaction was numerically highest for SuppCond-AutGrp (mean 2.97, SD 0.49) and lowest for AutCond-SuppGrp (mean 2.56, SD 0.59).

The paired sample  $t$  tests showed significant within-group differences for working alliance ( $t_{27}=3.20$ ;  $P=.003$ ) and satisfaction ( $t_{27}=2.86$ ;  $P=.008$ ) for the SuppGrp, with higher scores in the matching case than in the mismatch case, but not for the AutGrp. Independent sample  $t$  tests revealed significant differences for satisfaction in the AutCond between both groups ( $t_{53}=2.44$ ;  $P=.02$ ), with the AutGrp being more satisfied than the SuppGrp, but no significant differences were revealed for any other parameter.

Support satisfaction correlated with treatment satisfaction measured with the ZUF-8 for the congruent fit AutCond-AutGrp ( $r=0.60$ ;  $P=.001$ ) as well as for the incongruent fit AutCond-SuppGrp ( $r=0.69$ ;  $P<.001$ ) and, in a matching result,

support dissatisfaction negatively correlated with treatment satisfaction for the same fits (AutCond-AutGrp:  $r=-0.59$ ;  $P=.001$ ; AutCond-SuppGrp:  $r=-0.80$ ;  $P<.001$ ). Support satisfaction correlated with working alliance for the congruent fit AutCond-AutGrp ( $r=0.53$ ;  $P=.004$ ) and the incongruent fit AutCond-SuppGrp ( $r=0.58$ ;  $P=.001$ ), whereas support dissatisfaction negatively correlated with working alliance for AutCond-AutGrp ( $r=-0.50$ ;  $P=.008$ ) and AutCond-SuppGrp ( $r=-0.72$ ;  $P<.001$ ).

## Discussion

### Overview

Low adherence in self-guided internet-based interventions for depressive and other mental disorders is a significant issue and might be optimized by adapting the content and the context of the interventions to the personal needs and motives of the user, resulting in more engaging and less aversive experiences. In this nonclinical experimental study, the 2 motives *being autonomous* and *being supported* were used to test the hypothesis that a better motivational person-program fit is associated with less motivational incongruence and higher anticipated adherence. The formation of 2 groups was successful; the manipulation check suggested that a congruent person-program fit is associated with lower subjective motive violation and partly higher motive satisfaction (not for autonomy satisfaction) compared with a noncongruent fit, but the main hypotheses were only partially supported. This study represents a first attempt to experientially examine the adaptation of internet-based interventions to individuals' motives as a tool to increase adherence, working alliance, and satisfaction with the interventions. This is a novel way to personalize internet-based interventions, inspired by theories and findings from

motive-oriented therapeutic relationship building in face-to-face therapies [36,37,41,42].

### Manipulation

The formation of 2 groups based on the screening with the FAMOS was successful and revealed significant differences between the 2 groups in autonomy and dependence, with large effect sizes in independent measures. The manipulation (adapting the excerpts of the intervention and the form of the intervention to the 2 opposed motives) can be regarded as almost successful. The incongruent fits reported a significantly higher violation of the 2 motives than the congruent fits. Regarding the motive satisfaction, the congruent fit also resulted in a significantly higher support satisfaction (closest to *just right*) than the incongruent fit. However, this was not the case for autonomy satisfaction ( $P=.41$ ). The requirements for the experimental setup can therefore be considered as fulfilled for autonomy and support violation and support satisfaction but not for autonomy satisfaction. It may have been a decisive factor that the motives *being autonomous* and *being supported* were only explicitly captured by a questionnaire. It is conceivable that these constructs can be better assessed using implicit measures [60,61]. In addition, the autonomy-granting condition might have had a less strong effect than the support-granting condition, which is another explanation for the lack of differences in autonomy satisfaction.

### Main Results

Contrary to our expectations, there was no significant interaction between group and condition regarding the self-reported hypothetical expectation of the use of the *full version* of the program. Within-group differences for optimal anticipated adherence were significant in the SuppGrp. Thus, participants who described themselves as needing support reported that they would be more adherent to a program that matches their needs, corroborating the main hypotheses of this paper. However, surprisingly, this was not the case for the AutGrp. It therefore seems that the SuppGrp was more satisfied with the program overall (in contrast to the AutGrp) and that a congruent motive orientation for this group (SuppGrp) resulted in a higher anticipated adherence than an incongruent motive orientation. An explanation for these results could be that in the SuppCond, the texts directly referred to guidance in the form of personal therapeutic support (eg, “Personal guide will contact you regularly and ask how you are doing” and “Even if you do not contact me, I will check at least once a week how far you have progressed in the program and what you have entered in the worksheets to give you personal feedback with suggestions”). From a large base of studies, we know that guidance has a beneficial effect on adherence and effectiveness [2,8,15,62]. It is conceivable that the expectation of therapeutic guidance alone has a greater effect on anticipated adherence than more supportive language or wording alone.

Although for working alliance (measured using the WAI) neither a significant interaction effect nor a significant main effect for

group was shown, there was a significant main effect for condition: the SuppCond was accompanied by higher working alliance. Similarly, no significant interaction and group effects were found for satisfaction with the program (measured using the ZUF-8), whereas the main effect for condition was significant: the supportive program was expected to be more satisfying. As for anticipated adherence, within-group differences (match vs mismatch) for working alliance and satisfaction were significant for the SuppGrp, again corroborating one of our main hypotheses, but not for the AutGrp.

An explanation for the nonsignificant within-group results in the AutGrp might be that the manipulation for autonomy satisfaction did not completely work out. Therefore, the results must be considered with caution.

### Limitations

This study includes several limitations. An important limitation is that adherence to the internet-based intervention was not measured as true adherence on a behavioral level but as time in anticipated minutes working with the program. Thus, both the validity of the measures and the generalizability of the associated findings can be partially questioned. However, treatment expectations are known to have a major impact because they are among the most important predictors of outcomes [63,64]. Furthermore, the sample was not a clinical sample and was recruited mainly through student forums or psychology groups, again diminishing the generalizability of the findings. Particularly, the lack of distress and treatment motivation in this nonclinical sample is likely to have led to smaller effects (ie, because participants did not experience distress, they may not have felt it necessary to spend much time with the web-based program), which the study was not powered to detect. Taken together, the successful formation of groups apparently with matching and mismatching program versions was strong enough to induce motive satisfaction and violation, but the impact on more distal measures such as anticipated adherence might have been underestimated in this nonclinical sample. An additional shortcoming is that no control condition (evaluating a program with no motive adaptation) was included in the study, resulting in a limited validity of the results, and the motives were assessed only explicitly with the FAMOS and not implicitly.

### Conclusions

Our study should be seen as a first proof of concept from which programs could now be adapted and evaluated for other motives. Despite the inconclusive results regarding autonomy, the study gives reason to assume that motive orientation could have a positive impact on adherence, working alliance, and satisfaction in internet-based interventions for depression and other mental disorders. This should be investigated in future randomized controlled trials with clinical samples that allow assessment of the actual (and not only the anticipated) adherence, working alliance, and treatment satisfaction.

## Acknowledgments

The project was funded by a fellowship to SW by the Swiss National Science Foundation (grant number 177678).

## Authors' Contributions

LB wrote the article with support from SW, TB, and AB. SW and TB conceived the study. SW implemented the study and conducted the data collection with help from LB. All authors made substantial contributions to the final manuscript.

## Conflicts of Interest

None declared.

## References

1. Rommel A, Bretschneider J, Kroll L, Prütz F, Thom J. The utilization of psychiatric and psychotherapeutic services in Germany—individual determinants and regional differences. *Utilization of Health Services in Germany*. 2017. URL: [https://www.researchgate.net/publication/323357911\\_The\\_utilization\\_of\\_psychiatric\\_and\\_pschotherapeutic\\_services\\_in\\_Germany\\_-\\_individual\\_determinants\\_and\\_regional\\_differences](https://www.researchgate.net/publication/323357911_The_utilization_of_psychiatric_and_pschotherapeutic_services_in_Germany_-_individual_determinants_and_regional_differences) [accessed 2018-08-06]
2. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cogn Behav Ther* 2018 Jan;47(1):1-18. [doi: [10.1080/16506073.2017.1401115](https://doi.org/10.1080/16506073.2017.1401115)] [Medline: [29215315](https://pubmed.ncbi.nlm.nih.gov/29215315/)]
3. Ebert DD, Donkin L, Andersson G, Andrews G, Berger T, Carlbring P, et al. Does internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. *Psychol Med* 2016 Oct;46(13):2679-2693 [FREE Full text] [doi: [10.1017/S0033291716001562](https://doi.org/10.1017/S0033291716001562)] [Medline: [27649340](https://pubmed.ncbi.nlm.nih.gov/27649340/)]
4. Karyotaki E, Kemmeren L, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, et al. Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? An individual participant data meta-analysis. *Psychol Med* 2018 Nov;48(15):2456-2466 [FREE Full text] [doi: [10.1017/S0033291718000648](https://doi.org/10.1017/S0033291718000648)] [Medline: [29540243](https://pubmed.ncbi.nlm.nih.gov/29540243/)]
5. Lal S, Adair CE. E-mental health: a rapid review of the literature. *Psychiatr Serv* 2014 Jan 01;65(1):24-32. [doi: [10.1176/appi.ps.201300009](https://doi.org/10.1176/appi.ps.201300009)] [Medline: [24081188](https://pubmed.ncbi.nlm.nih.gov/24081188/)]
6. Richards D, Viganó N. Online counseling: a narrative and critical review of the literature. *J Clin Psychol* 2013 Sep;69(9):994-1011. [doi: [10.1002/jclp.21974](https://doi.org/10.1002/jclp.21974)] [Medline: [23630010](https://pubmed.ncbi.nlm.nih.gov/23630010/)]
7. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. *World Psychiatry* 2014 Oct;13(3):288-295 [FREE Full text] [doi: [10.1002/wps.20151](https://doi.org/10.1002/wps.20151)] [Medline: [25273302](https://pubmed.ncbi.nlm.nih.gov/25273302/)]
8. 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](https://doi.org/10.1016/j.invent.2014.08.003)]
9. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. *J Med Internet Res* 2017 Sep 15;19(9):e306 [FREE Full text] [doi: [10.2196/jmir.6588](https://doi.org/10.2196/jmir.6588)] [Medline: [28916506](https://pubmed.ncbi.nlm.nih.gov/28916506/)]
10. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. *World Psychiatry* 2014 Feb;13(1):4-11 [FREE Full text] [doi: [10.1002/wps.20083](https://doi.org/10.1002/wps.20083)] [Medline: [24497236](https://pubmed.ncbi.nlm.nih.gov/24497236/)]
11. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the internet: a systematic review of applications, clinical efficacy and cost-effectiveness. *Expert Rev Pharmacoecon Outcomes Res* 2012 Dec;12(6):745-764. [doi: [10.1586/erp.12.67](https://doi.org/10.1586/erp.12.67)] [Medline: [23252357](https://pubmed.ncbi.nlm.nih.gov/23252357/)]
12. Westermann S, Rüegg N, Lüdtke T, Moritz S, Berger T. Internet-based self-help for psychosis: findings from a randomized controlled trial. *J Consult Clin Psychol* 2020 Oct;88(10):937-950. [doi: [10.1037/ccp0000602](https://doi.org/10.1037/ccp0000602)] [Medline: [32790453](https://pubmed.ncbi.nlm.nih.gov/32790453/)]
13. Bücker L, Bierbrodt J, Hand I, Wittekind C, Moritz S. Effects of a depression-focused internet intervention in slot machine gamblers: a randomized controlled trial. *PLoS ONE* 2018 Jun 8;13(6):e0198859. [doi: [10.1371/journal.pone.0198859](https://doi.org/10.1371/journal.pone.0198859)]
14. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. *JAMA Psychiatry* 2017 Apr 01;74(4):351-359. [doi: [10.1001/jamapsychiatry.2017.0044](https://doi.org/10.1001/jamapsychiatry.2017.0044)] [Medline: [28241179](https://pubmed.ncbi.nlm.nih.gov/28241179/)]
15. Karyotaki E, Efthimiou O, Miguel C, Berman FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry* 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: [10.1001/jamapsychiatry.2020.4364](https://doi.org/10.1001/jamapsychiatry.2020.4364)] [Medline: [33471111](https://pubmed.ncbi.nlm.nih.gov/33471111/)]
16. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit J, et al. Adherence to internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. *PLoS One* 2014;9(7):e100674 [FREE Full text] [doi: [10.1371/journal.pone.0100674](https://doi.org/10.1371/journal.pone.0100674)] [Medline: [25029507](https://pubmed.ncbi.nlm.nih.gov/25029507/)]

17. Schmidt ID, Forand NR, Strunk DR. Predictors of dropout in internet-based cognitive behavioral therapy for depression. *Cognit Ther Res* 2019 Jun 16;43(3):620-630 [[FREE Full text](#)] [doi: [10.1007/s10608-018-9979-5](https://doi.org/10.1007/s10608-018-9979-5)] [Medline: [32879540](#)]
18. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. *Psychol Med* 2015 Oct;45(13):2717-2726. [doi: [10.1017/S0033291715000665](https://doi.org/10.1017/S0033291715000665)] [Medline: [25881626](#)]
19. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. *J Med Internet Res* 2009 Apr 24;11(2):e13 [[FREE Full text](#)] [doi: [10.2196/jmir.1194](https://doi.org/10.2196/jmir.1194)] [Medline: [19403466](#)]
20. Bücker L, Schnakenberg P, Karyotaki E, Moritz S, Westermann S. Diminishing effects after recurrent use of self-guided internet-based interventions in depression: randomized controlled trial. *J Med Internet Res* 2019 Oct 02;21(10):e14240 [[FREE Full text](#)] [doi: [10.2196/14240](https://doi.org/10.2196/14240)] [Medline: [31579014](#)]
21. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. *J Med Internet Res* 2011 Aug 05;13(3):e52 [[FREE Full text](#)] [doi: [10.2196/jmir.1772](https://doi.org/10.2196/jmir.1772)] [Medline: [21821503](#)]
22. Lütke T, Pult LK, Schröder J, Moritz S, Bücker L. A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. *Psychiatry Res* 2018 Nov;269:753-762. [doi: [10.1016/j.psychres.2018.08.113](https://doi.org/10.1016/j.psychres.2018.08.113)] [Medline: [30273901](#)]
23. Titov N, Dear B, Johnston L, Lorian C, Zou J, Wootton B, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. *PLoS One* 2013;8(7):e62873 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0062873](https://doi.org/10.1371/journal.pone.0062873)] [Medline: [23843932](#)]
24. Fisher AJ, Boswell JF. Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment* 2016 Aug 14;23(4):496-506. [doi: [10.1177/1073191116638735](https://doi.org/10.1177/1073191116638735)] [Medline: [26975466](#)]
25. Swift JK, Callahan JL, Cooper M, Parkin SR. The impact of accommodating client preference in psychotherapy: a meta-analysis. *J Clin Psychol* 2018 Nov 08;74(11):1924-1937. [doi: [10.1002/jclp.22680](https://doi.org/10.1002/jclp.22680)] [Medline: [30091140](#)]
26. Beutler LE, Harwood TM, Michelson A, Song X, Holman J. Resistance/reactance level. *J Clin Psychol* 2011 Feb 24;67(2):133-142. [doi: [10.1002/jclp.20753](https://doi.org/10.1002/jclp.20753)] [Medline: [21108314](#)]
27. Beutler L, Edwards C, Someah K. Adapting psychotherapy to patient reactance level: a meta-analytic review. *J Clin Psychol* 2018 Nov;74(11):1952-1963. [doi: [10.1002/jclp.22682](https://doi.org/10.1002/jclp.22682)] [Medline: [30334254](#)]
28. Andrews G, Williams AD. Internet psychotherapy and the future of personalized treatment. *Depress Anxiety* 2014 Nov 18;31(11):912-915. [doi: [10.1002/da.22302](https://doi.org/10.1002/da.22302)] [Medline: [25407580](#)]
29. Berger T, Boettcher J, Caspar F. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. *Psychotherapy (Chic)* 2014 Jun;51(2):207-219. [doi: [10.1037/a0032527](https://doi.org/10.1037/a0032527)] [Medline: [24041199](#)]
30. Carlbring P, Maurin L, Törngren C, Linna E, Eriksson T, Sparthar E, et al. Individually-tailored, internet-based treatment for anxiety disorders: a randomized controlled trial. *Behav Res Ther* 2011 Jan;49(1):18-24. [doi: [10.1016/j.brat.2010.10.002](https://doi.org/10.1016/j.brat.2010.10.002)] [Medline: [21047620](#)]
31. Andersson G, Estling F, Jakobsson E, Cuijpers P, Carlbring P. Can the patient decide which modules to endorse? An open trial of tailored internet treatment of anxiety disorders. *Cogn Behav Ther* 2011 Mar;40(1):57-64. [doi: [10.1080/16506073.2010.529457](https://doi.org/10.1080/16506073.2010.529457)] [Medline: [21337215](#)]
32. Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, et al. Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. *PLoS One* 2012;7(5):e36905 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0036905](https://doi.org/10.1371/journal.pone.0036905)] [Medline: [22615841](#)]
33. Silfvernagel K, Westlinder A, Andersson S, Bergman K, Diaz Hernandez R, Fallhagen L, et al. Individually tailored internet-based cognitive behaviour therapy for older adults with anxiety and depression: a randomised controlled trial. *Cogn Behav Ther* 2018 Jul 25;47(4):286-300. [doi: [10.1080/16506073.2017.1388276](https://doi.org/10.1080/16506073.2017.1388276)] [Medline: [29068266](#)]
34. Moritz S, Stepulovs O, Schröder J, Hottenrott B, Meyer B, Hauschildt M. Is the whole less than the sum of its parts? Full versus individually adapted metacognitive self-help for obsessive-compulsive disorder: A randomized controlled trial. *J Obsessive Compulsive Relate Disord* 2016 Apr;9:107-115. [doi: [10.1016/j.jocrd.2016.04.001](https://doi.org/10.1016/j.jocrd.2016.04.001)]
35. Westermann S, Moritz S, Berger T. Internet- und mobilbasierte Interventionen bei Schizophrenie. *Verhaltenstherapie* 2017 Aug 18;27(3):181-189. [doi: [10.1159/000479335](https://doi.org/10.1159/000479335)]
36. Rosner R. *Psychologische Therapie (Psychological Therapy)*. *Psychother Res* 2010 Apr 23;11(4):484-486. [doi: [10.1080/713664068](https://doi.org/10.1080/713664068)]
37. Grawe K. *Psychological Therapy*. Göttingen, Germany: Hogrefe Publishing; 2004.
38. Grosse Holtforth M, Wilm K, Beyermann S, Rhode A, Trost S, Steyer R. Differential change in integrative psychotherapy: a re-analysis of a change-factor based RCT in a naturalistic setting. *Psychother Res* 2011 Nov;21(6):631-643. [doi: [10.1080/10503307.2011.602749](https://doi.org/10.1080/10503307.2011.602749)] [Medline: [21793690](#)]
39. Grosse Holtforth M, Grawe K. Inkongruenz und Fallkonzeption in der Psychologischen Therapie. *Verhaltenstherapie Psychosoziale Praxis* 2004;36(1):9-21.
40. Caspar F. Plan analysis. In: *Handbook of Psychotherapy Case Formulations*. New York: Guilford Press; 2007.

41. Caspar F. *Understand Relationships and Problems An Introduction to Psychotherapeutic Plan Analysis*. Goettingen Germany: Hogrefe Verlag GmbH & Co; 1996.
42. Grosse Holtforth M, Castonguay LG. Relationship and techniques in cognitive-behavioral therapy--a motivational approach. *Psychother Theory Res Pract Train* 2005;42(4):443-455 [FREE Full text] [doi: [10.1037/0033-3204.42.4.443](https://doi.org/10.1037/0033-3204.42.4.443)]
43. Grosse Holtforth M, Grawe K. Fragebogen zur Analyse Motivationaler Schemata (FAMOS). *Zeitschrift für Klinische Psychologie und Psychotherapie* 2000 Jul;29(3):170-179. [doi: [10.1026//0084-5345.29.3.170](https://doi.org/10.1026//0084-5345.29.3.170)]
44. Faul F, Erdfelder E, Lang A, Buchner A. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007 May;39(2):175-191. [doi: [10.3758/bf03193146](https://doi.org/10.3758/bf03193146)] [Medline: [17695343](https://pubmed.ncbi.nlm.nih.gov/17695343/)]
45. Horvath AO, Greenberg LS. Development and validation of the working alliance inventory. *J Counsel Psychol* 1989 Apr;36(2):223-233. [doi: [10.1037/0022-0167.36.2.223](https://doi.org/10.1037/0022-0167.36.2.223)]
46. Hatcher RL, Gillaspay JA. Development and validation of a revised short version of the working alliance inventory. *Psychother Res* 2006 Jan;16(1):12-25. [doi: [10.1080/10503300500352500](https://doi.org/10.1080/10503300500352500)]
47. Wilmers F, Munder T, Leonhart R, Herzog T, Plassmann R, Linster HW, et al. Die deutschsprachige Version des Working Alliance Inventory - short revised (WAI-SR) - Ein schulentübergreifendes, ökonomisches und empirisch validiertes Instrument zur Erfassung der therapeutischen Allianz. *Klinische Diagnostik und Evaluation* 2008;1(3):343-358.
48. Munder T, Wilmers F, Leonhart R, Linster H, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clin Psychol Psychother* 2010;17(3):231-239. [doi: [10.1002/cpp.658](https://doi.org/10.1002/cpp.658)] [Medline: [20013760](https://pubmed.ncbi.nlm.nih.gov/20013760/)]
49. Schmidt J, Lamprecht F, Wittmann W. Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen [Satisfaction with inpatient care: Development of a questionnaire and first validity assessments]. *PPmP Psychotherapie Psychosomatik Medizinische Psychologie* 1989;39(7):248-255.
50. Kriz D, Nübling R, Steffanowski A, Wittmann W, Schmidt J. Patientenzufriedenheit in der stationären Rehabilitation: psychometrische Reanalyse des ZUF-8 auf der Basis multizentrischer Stichproben verschiedener Indikation. *Zeitschrift für Medizinische Psychologie* 2008;17(2-3):67-79.
51. Viglione DJ, Lovette GJ, Gottlieb R, Friedberg R. Depressive experiences questionnaire: an empirical exploration of the underlying theory. *J Pers Assess* 1995 Aug;65(1):91-99. [doi: [10.1207/s15327752jpa6501\\_7](https://doi.org/10.1207/s15327752jpa6501_7)] [Medline: [16367648](https://pubmed.ncbi.nlm.nih.gov/16367648/)]
52. Blatt SJ, D'Afflitti JP, Quinlan DM. Experiences of depression in normal young adults. *J Abnormal Psychol* 1976 Aug;85(4):383-389. [doi: [10.1037/0021-843x.85.4.383](https://doi.org/10.1037/0021-843x.85.4.383)]
53. Campos R, Holden R, Baleizão C, Caçador B, Fragata A. Self-criticism, neediness, and distress in the prediction of suicide ideation: results from cross-sectional and longitudinal studies. *J Psychol* 2018 May 19;152(4):237-255. [doi: [10.1080/00223980.2018.1446895](https://doi.org/10.1080/00223980.2018.1446895)] [Medline: [29667539](https://pubmed.ncbi.nlm.nih.gov/29667539/)]
54. Weinstein N, Przybylski AK, Ryan RM. The index of autonomous functioning: development of a scale of human autonomy. *J Res Personal* 2012 Aug;46(4):397-413. [doi: [10.1016/j.jrp.2012.03.007](https://doi.org/10.1016/j.jrp.2012.03.007)]
55. Schröder J, Sautier L, Kriston L, Berger T, Meyer B, Späth C, et al. Development of a questionnaire measuring attitudes towards Psychological Online Interventions-the APOI. *J Affect Disord* 2015 Nov 15;187:136-141. [doi: [10.1016/j.jad.2015.08.044](https://doi.org/10.1016/j.jad.2015.08.044)] [Medline: [26331687](https://pubmed.ncbi.nlm.nih.gov/26331687/)]
56. Grosse Holtforth M, Grawe K. Der Inkongruenzfragebogen (INK). *Zeitschrift für Klinische Psychologie und Psychotherapie* 2003 Oct;32(4):315-323. [doi: [10.1026/0084-5345.32.4.315](https://doi.org/10.1026/0084-5345.32.4.315)]
57. 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](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
58. Smarr KL, Keefer AL. Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care Res (Hoboken)* 2011 Nov 07;63 Suppl 11(S11):S454-S466 [FREE Full text] [doi: [10.1002/acr.20556](https://doi.org/10.1002/acr.20556)] [Medline: [22588766](https://pubmed.ncbi.nlm.nih.gov/22588766/)]
59. Lehl S, Triebig G, Fischer B. Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurol Scand* 1995 May;91(5):335-345. [doi: [10.1111/j.1600-0404.1995.tb07018.x](https://doi.org/10.1111/j.1600-0404.1995.tb07018.x)] [Medline: [7639062](https://pubmed.ncbi.nlm.nih.gov/7639062/)]
60. Baum IR, Baumann N. Arousing autonomy: a valid assessment of the implicit autonomy motive. *Personal Individ Diff* 2021 Jan;168:110362. [doi: [10.1016/j.paid.2020.110362](https://doi.org/10.1016/j.paid.2020.110362)]
61. Rawolle M, Schultheiss M, Schultheiss OC. Relationships between implicit motives, self-attributed motives, and personal goal commitments. *Front Psychol* 2013;4:923 [FREE Full text] [doi: [10.3389/fpsyg.2013.00923](https://doi.org/10.3389/fpsyg.2013.00923)] [Medline: [24367351](https://pubmed.ncbi.nlm.nih.gov/24367351/)]
62. Bur OT, Krieger T, Moritz S, Klein JP, Berger T. Optimizing the context of support of web-based self-help in individuals with mild to moderate depressive symptoms: a randomized full factorial trial. *Behav Res Ther* 2022 May;152:104070 [FREE Full text] [doi: [10.1016/j.brat.2022.104070](https://doi.org/10.1016/j.brat.2022.104070)] [Medline: [35306266](https://pubmed.ncbi.nlm.nih.gov/35306266/)]
63. Rief W, Hofmann S, Nestoriuc Y. The power of expectation - Understanding the placebo and nocebo phenomenon. *Soc Personal Psychol Compass* 2008;2(4):1624-1637. [doi: [10.1111/j.1751-9004.2008.00121.x](https://doi.org/10.1111/j.1751-9004.2008.00121.x)]
64. Rief W, Anna Glombiewski J. The role of expectations in mental disorders and their treatment. *World Psychiatry* 2017 Jun 12;16(2):210-211 [FREE Full text] [doi: [10.1002/wps.20427](https://doi.org/10.1002/wps.20427)] [Medline: [28498580](https://pubmed.ncbi.nlm.nih.gov/28498580/)]

## Abbreviations

**APOI:** Attitudes Toward Psychological Online Interventions  
**AutCond:** autonomy condition  
**AutGrp:** autonomy group  
**FAMOS:** Fragebogen zur Analyse Motivationaler Schemata  
**IAF:** Index of Autonomous Functioning  
**K-INK:** Kurzversion des Inkongruenzfragebogen  
**MWT-B:** Multiple-Choice Vocabulary Intelligence Test  
**PHQ-9:** Patient Health Questionnaire-9, depression module  
**SuppCond:** support condition  
**SuppGrp:** support group  
**TDEQ-12:** Theoretical Depressive Experiences Questionnaire, short version  
**WAI:** Working Alliance Inventory  
**ZUF-8:** Zufriedenheitsfragebogen

*Edited by A Mavragani; submitted 14.02.22; peer-reviewed by D Erbe, N Forand; comments to author 17.05.22; revised version received 07.06.22; accepted 09.06.22; published 13.09.22.*

*Please cite as:*

*Bücker L, Berger T, Bruhns A, Westermann S*

*Motive-Oriented, Personalized, Internet-Based Interventions for Depression: Nonclinical Experimental Study*

*JMIR Form Res 2022;6(9):e37287*

*URL: <https://formative.jmir.org/2022/9/e37287>*

*doi: [10.2196/37287](https://doi.org/10.2196/37287)*

*PMID: [36098989](https://pubmed.ncbi.nlm.nih.gov/36098989/)*

©Lara Bücker, Thomas Berger, Alina Bruhns, Stefan Westermann. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 13.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Development of a Mobile App for Self-Care Against COVID-19 Using the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) Model: Methodological Study

Hamid Reza Saeidnia<sup>1\*</sup>, MSc; Marcin Kozak<sup>2\*</sup>, PhD; Marcel Ausloos<sup>3\*</sup>, PhD; Claudiu Herteliu<sup>4</sup>, PhD; Zahra Mohammadzadeh<sup>5\*</sup>, PhD; Ali Ghorbi<sup>6</sup>, MSc; Mehrdad Karajizadeh<sup>7</sup>, PhD; Mohammad Hassanzadeh<sup>1\*</sup>, PhD

<sup>1</sup>Department of Knowledge and Information Science, Tarbiat Modares University, Tehran, Iran

<sup>2</sup>Department of Media, Journalism and Social Communication, University of Information Technology and Management in Rzeszow, Rzeszow, Poland

<sup>3</sup>School of Business, University of Leicester, Brookfield, Leicester, United Kingdom

<sup>4</sup>Department of Statistics and Econometrics, Bucharest University of Economic Studies, Bucharest, Romania

<sup>5</sup>Department of Health Information Management and Technology, School of Allied Medical Sciences, Kashan University of Medical Sciences, Kashan, Iran

<sup>6</sup>Department of Knowledge and Information Science, University of Tehran, Tehran, Iran

<sup>7</sup>Trauma Research Center, Shahid Rajaee (Emtiaz) Trauma Hospital, Shiraz University of Medical Sciences, Shiraz, Fars, Iran

\*these authors contributed equally

**Corresponding Author:**

Hamid Reza Saeidnia, MSc

Department of Knowledge and Information Science

Tarbiat Modares University

Tehran-Ashrafi Esfahani Highway- After Paul Hemmat

Tehran, 6931753865

Iran

Phone: 98 09394464224

Email: [hamidrezaednia@gmail.com](mailto:hamidrezaednia@gmail.com)

## Abstract

**Background:** Mobile apps have been shown to play an important role in the management, care, and prevention of infectious diseases. Thus, skills for self-care—one of the most effective ways to prevent illness—can be improved through mobile health apps.

**Objective:** This study aimed to design, develop, and evaluate an educational mobile-based self-care app in order to help the self-prevention of COVID-19 in underdeveloped countries. We intended the app to be easy to use, quick, and inexpensive.

**Methods:** In 2020 and 2021, we conducted a methodological study. Using the ADDIE (analysis, design, development, implementation, and evaluation) educational model, we developed a self-care management mobile app. According to the ADDIE model, an effective training and performance support tool is built through the 5 phases that comprise its name. There were 27 participants who conducted 2 evaluations of the mobile app's usability and impact using the mobile health app usability and self-care inventory scales. The study design included pre- and posttesting.

**Results:** An Android app called MyShield was developed. The results of pre- and posttests showed that on a scale from 0 to 5, MyShield scored a performance average of 4.17 in the physical health dimension and an average of 3.88 in the mental well-being dimension, thereby showing positive effects on self-care skills. MyShield scored highly on the “interface and satisfaction,” “ease of use,” and “usefulness” components.

**Conclusions:** MyShield facilitates learning self-care skills at home, even during quarantine, increasing acquisition of information. Given its low development cost and the ADDIE educational design on which it is based, the app can be helpful in underdeveloped countries. Thus, low-income countries—often lacking other tools—can use the app as an effective tool for fighting COVID-19, if it becomes a standard mobile app recommended by the government.

(JMIR Form Res 2022;6(9):e39718) doi:[10.2196/39718](https://doi.org/10.2196/39718)

**KEYWORDS**

self-care; mobile app; ADDIE model; COVID-19; underdeveloped countries

**Introduction**

In 2019 and 2020, COVID-19 attracted international attention as a serious threat to public health [1,2]. Thereafter, the virus spread rapidly around the world. On March 11, 2020, the World Health Organization declared the epidemic to be a pandemic [3]. The world then faced a situation that seemed impossible: without war, the global population was in danger; COVID fatalities mounted, and nothing seemed to make people safe. What remained was hope, and small steps everyone could take, such as isolation and good hygiene. This is where simple tools can help, and some examples of such tools are those related to digital technology.

Higher-income and further-developed countries such as Japan, the United States, and South Korea indeed started using digital technology to fight COVID-19, including such methods as contact tracing, the internet of things, big-data analytics, and artificial intelligence [4]. In other, similarly developed countries, such as Canada and Australia, mobile apps (eg, COVID Alert and COVIDSafe) were used to inform citizens about COVID-19 and its prevention [5-7]. These tools appeared to be working well: with a relatively low usage cost to both the governmental and individual user, their effects were quite noticeable.

Nonetheless, most people in lower-income and underdeveloped countries lack access to complex digital technology, as these countries' levels of information and communications technology infrastructure are much lower than those of more developed countries [8,9]. Several studies have recently indicated that mortality, hospitalization rate, and complications due to COVID-19 in underdeveloped countries have been higher than in other countries [8-10]. It is conventionally agreed upon that due to demographic features, weak health care systems, inequality in the economic and political sectors, corruption, and other sociocultural characteristics, many underdeveloped countries have lower levels of protection against the spread of COVID-19 and its subsequent effects [8,9,11]. In addition to these factors, a lack of social distancing and self-care knowledge contributed to an increased incidence of COVID-19 and its variations, highlighted by a mutation first reported in India commonly known as the Delta variant [12]. All viruses have variants, but the coronavirus' new mutations make it unpredictable and even more difficult to fight.

For the above reasons, learning self-care is important [13,14]. For lower-income and socially unequal communities, self-care is considered to be the main way of responding to the disease's spread and severity, mainly because of self-care's low cost [15]. In self-care, individuals use their own knowledge and abilities to maintain and improve their health [16]. It is a challenging skill that requires time and energy, because its implementation depends on internal (ie, cognitive, physical, emotional, and behavioral) and external (ie, environmental, political, and societal) factors [14]. Meanwhile, self-care has become a new strategy for managing and preventing chronic and infectious diseases, leading to increased energy, more positive emotions,

reduced stress, improved health and well-being, and increased self-confidence [17,18]. Self-care is where mobile apps enter the scene, with their capabilities to help people learn self-care tasks.

COVID-19-related smartphone apps offer useful capabilities and functionalities, such as tracking, follow-up, and prevention [18-20]. Furthermore, the growing popularity of mobile software and its relatively low cost have led to the emergence of mobile self-care systems [21]. Not only are such systems relatively cheap and easy to develop, but they also help to manage disease prevention. It is no wonder that in recent years, mobile software has been applied in various medical and health fields [18-21]. In light of these advancements, current digital technologies are highly productive despite their low cost, and they can greatly support health management in lower-income and underdeveloped countries.

While a basic mobile app can be built solely by a person or persons with sufficient development skills, such an app is unlikely to become a useful, popular app; it has been built using only the knowledge and skills of the developer or developers, without input or feedback from specialized experts. Knowing this, we decided to design our app with deeper knowledge in the areas of mobile app design, development, and COVID-19 health information. This involved conducting a study in which we would ask various experts—health practitioners and mobile technology professionals—for their help; in this way, the resulting app would be designed using much deeper and broader knowledge than a single author or team could provide.

We also decided to use the ADDIE (analysis, design, development, implementation, and evaluation) education model, which is used to build educational systems. While the app we wanted to create was not an educational system per se, it did aim to teach, and we wanted it to teach effectively. We observed a strong correlation between our concept of a mobile app and educational systems created using the ADDIE model. Why did we decide to use this model to create a mobile app? We believed that an app that was designed using an educational model had a greater chance to teach its users effectively than an app designed without an educational methodology.

Therefore, in this study, we aimed to design, develop, and implement a standalone mobile app for training self-care skills based on the ADDIE educational model in order to improve self-care and prevent COVID-19, particularly in underdeveloped countries. We evaluated various facets of the app, all directly or indirectly related to its main goal: to help its users enhance their self-care skills in the context of COVID-19.

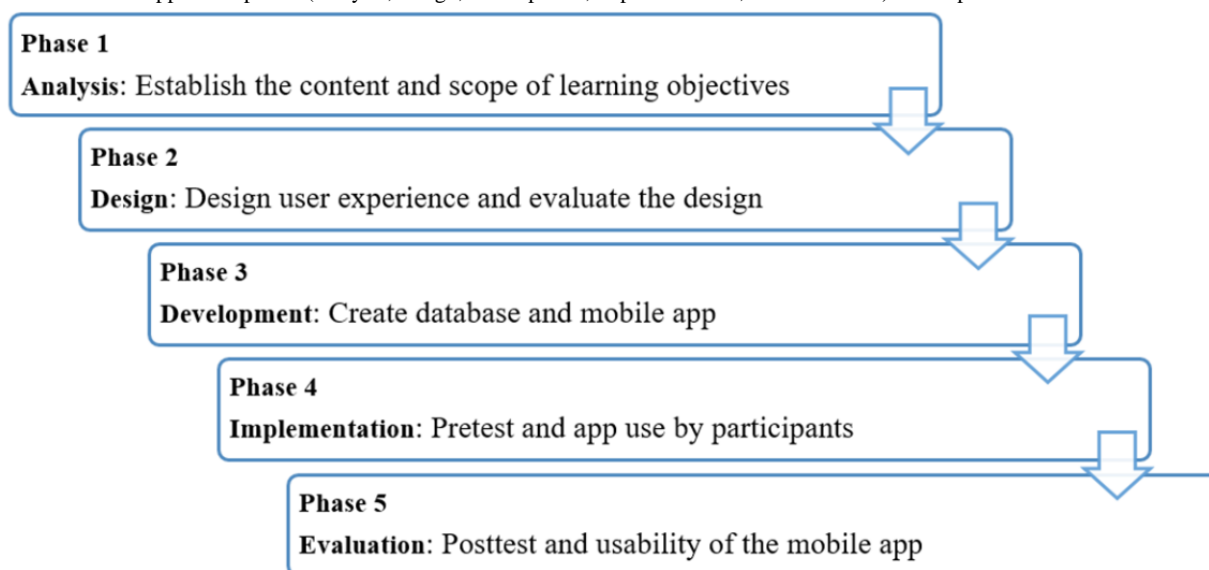
**Methods****Overview**

The present methodological study used the ADDIE educational design model to develop an educational mobile app for self-care with the purpose of supporting self-prevention of COVID-19 [22]. ADDIE gives education professionals and designers the

ability to lay the groundwork for principled and effective training [23]. This model can be utilized for both traditional and electronic learning [24]. ADDIE consists of the 5 phases

the acronym was built upon: analysis, design, development, implementation, and evaluation (Figure 1) [23].

**Figure 1.** Educational app development (analysis, design, development, implementation, and evaluation) model phases.



### Analysis Phase

During the analysis phase, we sought to determine the appropriate practical contents of the mobile app and categorize them. For this purpose, from October 1 to 15, 2020, we interviewed 3 infectious disease experts and 2 nursing specialists. Because of COVID-19 conditions, the interviews were remote and used the Zoom meeting software. We used qualitative methods to analyze and categorize the resulting content requirements for a self-care mobile app [18,25]. In order to make the app pleasant to use, we used interesting content presented in various formats (eg, written text, images, infographics, and videos).

### Design Phase

In the design phase, we focused on optimizing the user experience of the designed mobile app and its feature list; we did so through online workshops with a group of experts in the fields of health information technology, medical informatics, user experience design, and Android mobile app development. The experts were enrolled via requests on social networks (eg, Instagram, Facebook, and LinkedIn) sent from October 20 to 28, 2020; we tried to enroll as many experts as possible.

According to the literature, we required between 14 and 28 experts [26]. Ultimately, 14 people agreed to participate in the design phase. In this phase, we used the Balsamiq Mockups software (Balsamiq Studios, LLC) to draw the user experience flow, and Figma (Figma, Inc) to prototype the user experience. Once the experts accepted an alpha version of the app's prototype, we evaluated it using the heuristic method, in order to determine those elements of the app that needed improvement before the development phase.

We asked 5 experts in the field of user experience design and development of mobile health (mHealth) apps for assistance. They ranked the prototype using Jacob Nielsen's 10 general

principles of severity for evaluating interactive design, which use a 5-point scale [27,28]: 0 ("I do not agree that this is a usability issue"), 1 ("Not to be fixed unless extra time is available for the project"), 2 ("This should be fixed, but at the lowest priority"), 3 ("Must be fixed, so it should be given high priority"), and 4 ("This must be fixed before the product can be released").

### Development Phase

The results of the analysis and design phases provided the knowledge to be used in the development phase, in which we developed a database using My Structured Query Language (MySQL) and a mobile app, which we called "MyShield," using Android Studio. The development phase took place from November 1 to 8, 2020.

### Implementation Phase

During this stage, we recruited ordinary people (ie, nonexperts) for our study through purposive sampling. Interested candidates contacted us through online registration after seeing the recruitment posts on social networks (including Instagram, Facebook, and LinkedIn). Twenty-seven participants registered; this was within the intended range of 25 to 30 subjects, as used in other, similar studies [26,29]. The participants were aged 18 to 50 years and were required to own a smartphone running Android (minimum version 4.1).

We used an online workshop to introduce the research team and project aims. During the workshop, we explained the ADDIE design process to the participants and their involvement. As a pretest, conducted before using the app, the participants were required to complete a self-care inventory (SCI) questionnaire. This questionnaire includes 30 items and 6 dimensions [30]. The questions were answered using a 5-point scale (almost never, occasionally, half of the time, fairly often, and almost always). In this questionnaire, each dimension included 5 questions; the respondents scored each question using this scale.

Then, we calculated each respondent's score by adding the points from all 6 dimensions; the result shows the person's level of self-care ([Multimedia Appendix 1](#)). The final score ranged from 30 to 150, with the following interpretations [30]: 120 or above, the person has personal well-being and serenity; 91 to 119, the person has some control of a good system of self-care; 50 to 90, the person is struggling and could use some assistance in developing a stronger self-care system; and under 50, the person experiences some serious difficulties in the area of self-care.

Once the participants completed the questionnaires, they started using the MyShield app, from November 10 to December 25, 2020 (45 days). During these 45 days, we asked the participants to use the app every day for as long as they wanted, but for at least 20 minutes per day; the time of day did not matter. The app contains an initial guide (presented as a showcase); we also created a WhatsApp group where the participants could ask us any questions related to the study and the app. The initial guide feature displayed the guidelines on how to work with the app after the user's first encounter with it. The app also contained an "About Us" section, where users could find information about the research, the app design team, and how to contact the team. The app allowed users to update to a newer version in the case of any resolved technical issues.

### Evaluation Phase

After the implementation phase ended on December 26, 2020, we used the standard standalone MAUQ (Mobile Health App Usability Questionnaire) to evaluate the usability of the app. This questionnaire has 3 factors: ease of use (5 items), interface and satisfaction (7 items), and usefulness (6 items). It is based on a 7-item Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) [31]. In the end, we converted each factor into a percentage between 0 and 100 based on each item's mean. We reused the self-care inventory questionnaire [30] in a posttest to analyze how the app influenced the participants' self-care skills; we did so by comparing these results with the pretest conducted during the implementation phase. Both questionnaires were designed with Google Forms, whose web addresses were sent to the WhatsApp group. The data were analyzed using SPSS software version 22 (SPSS Inc).

### Ethics Approval

The participants in the pretest and posttest surveys provided informed consent. Tarbiat Modares University Ethics Committee (IR. MODARES. REC.1399.142) approved the entire study protocol.

## Results

### Design and Evaluation

In the analysis phase, meetings with the experts helped us find the optimal content of the MyShield app, including the main menu. The participants agreed that MyShield should cover the following 5 topics: self-motivation, daily life management, personal hygiene, healthy eating, and exercise.

We determined that self-motivation content should teach the users how to think positively, motivate themselves, and utilize

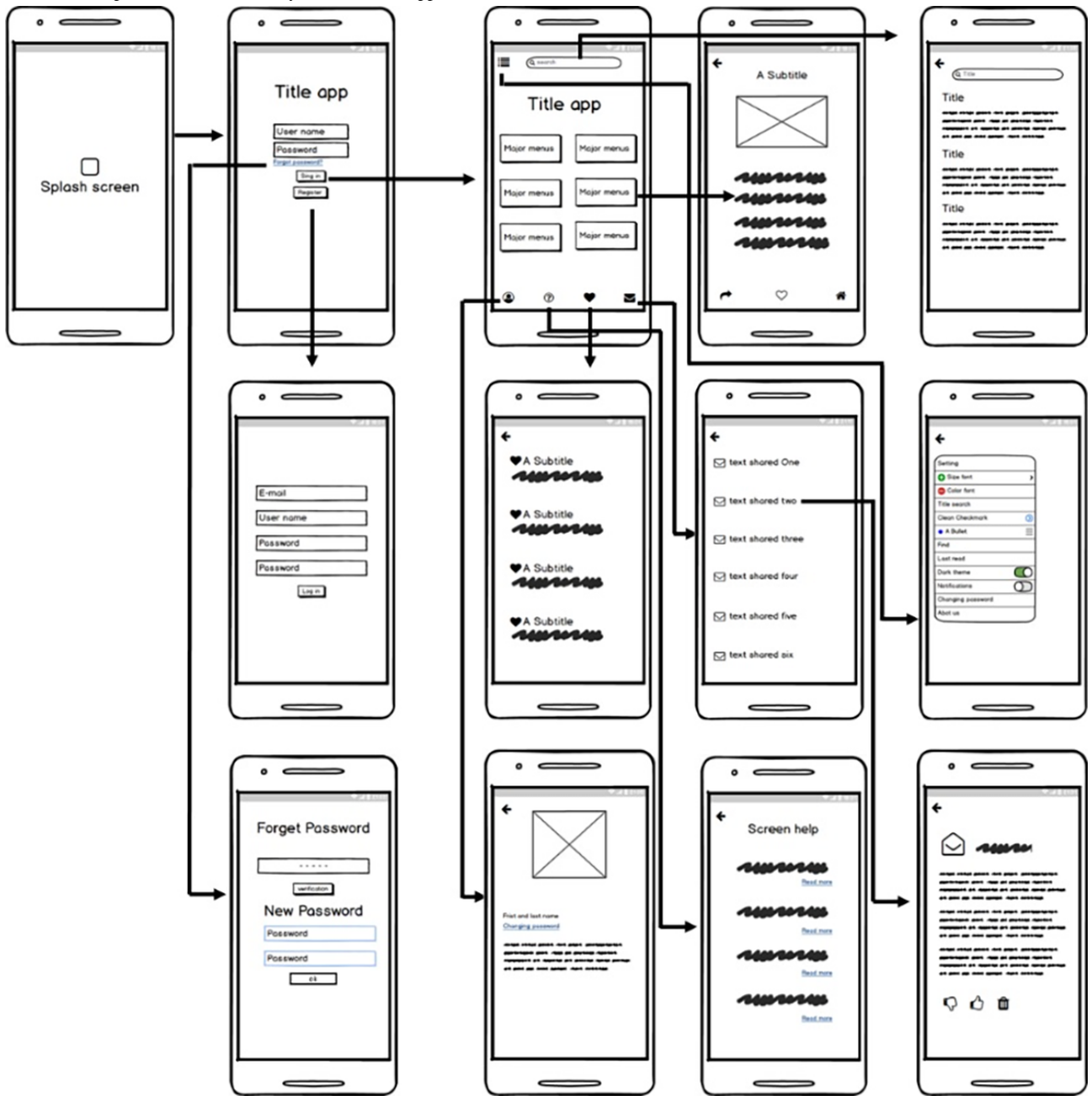
their intrinsic motivations. In matters related to daily life management, the app should include strategies on how to manage stress, fear, and anger. Education is also essential in other areas, such as managing life during personal quarantine and social activities. The personal hygiene aspect should deal with education related to masking and observing social distancing, washing hands, and use of disinfectants. Healthy eating education should include information on useful supplements, beverages, fatty foods, proteins, dairy products, nutrition programs, and controlling or avoiding smoking. Exercise education should cover such topics as exercising at home, conditions for exercising outdoors, professional exercise, the intensity of exercise, and exercise in self-quarantine.

Online workshops conducted with the 14 experts were utilized to determine the practical features; moreover, the user experience decided during the design phase helped us to determine the practical design of MyShield, as follows. The app starts with an initial opening screen. Next, the registration page appears. Here, the user must provide a few pieces of personal information (email, user name, and password). An introduction slide appears on the initial launch of the app. Three slides appear to be ideal for introducing the app. We found that in order to familiarize the user with the app, the showcase should appear after the introduction slide. After this, the user sees the main menu. It has two action bars: one, at the top of the page, displays the search feature of the entire app; while the other, at the bottom of the page, contains buttons, messages, help, settings, and profiles. On the left-hand side, there is a hamburger button (3 horizontal lines) which, when tapped, leads to the app's settings. The settings include fonts, title searches, option selection and clearing, find, last read, light or dark theme, notification, changing passwords, and an "About Us" option. The content had a share button, which should be checked as marked. In addition, the users had an option of liking or disliking the content. [Figure 2](#) shows the user experience flow. For example, tapping the "sign in" button opens the Main Page screen. Thus, this figure is both a user experience flow and a wireframe, which is an illustrated guide that shows the schematic and general framework of the MyShield app.

During the heuristic evaluation of the app's prototype, these features each earned 0 to 2 points, meaning that the prototype received a rejection or an acceptable score. [Table 1](#) presents the results:  $\Sigma A$  represents the total number of points for each item from Jacob Nielsen's 10 general principles;  $\Sigma B$  is the same as  $\Sigma A$  with duplicated points eliminated. The total for  $\Sigma A$  was 45 and the total for  $\Sigma B$  was 23 ([Table 1](#)).

In the development phase, functional requirements, user interface screens, and software database designs were created. The interface design was a process that involved installing specialized software on smartphones and connecting them to the internet, so that users could register their accounts in the app. Thus, the users were registered on the server and their registration IDs were stored on the app server. MySQL was used to design and develop the database, and Android Studio was used to develop the app. Eventually, after thorough consultation with infectious disease specialists and experienced nurses, a beta version of MyShield was developed and released ([Figure 3](#)).

Figure 2. User experience flow of the MyShield mobile app.



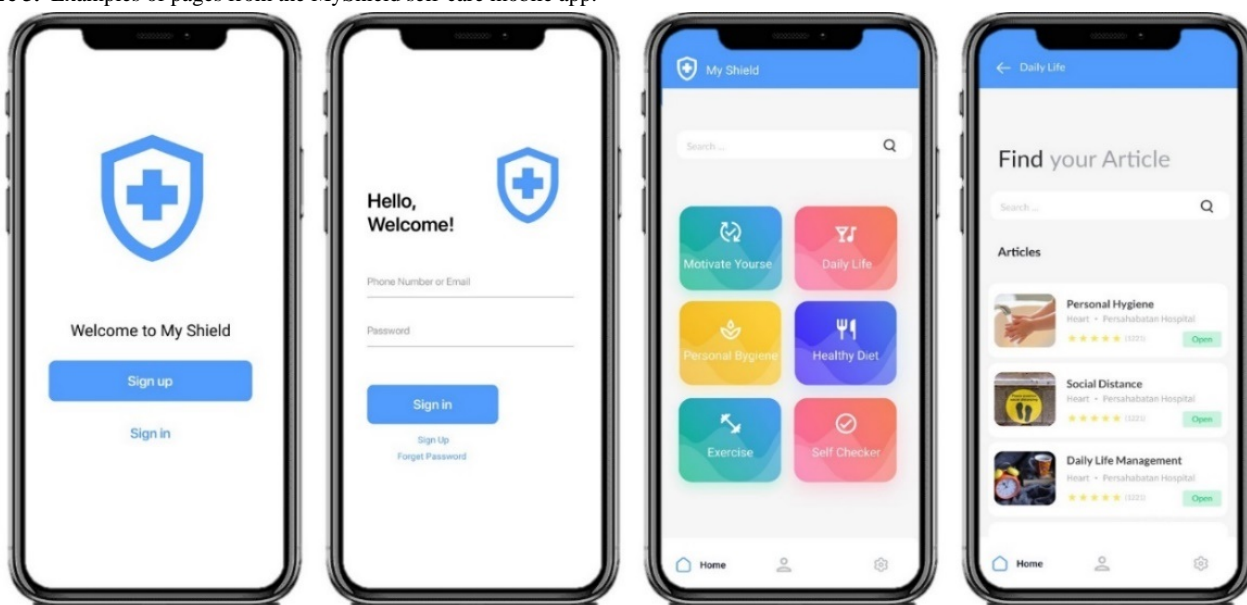
**Table 1.** MyShield’s prototype scoring according to a heuristic evaluation based on Jacob Nielsen’s 10 general principles.

Heuristic principle	ΣA <sup>a</sup>	ΣB <sup>b</sup>	Expert identification number				
			1	2	3	4	5
Visibility of system status	4	3	0	1	2	1	0
Match between system and the real world	2	2	0	0	0	0	2
User control and freedom	4	3	1	1	0	2	0
Consistency and standards	2	1	0	1	0	0	1
Error prevention	6	3	2	2	1	1	0
Recognition rather than recall	3	1	0	0	1	1	1
Flexibility and efficiency of use	9	3	2	2	2	2	1
Aesthetic and minimalist design	6	3	1	0	1	2	2
Error identification, diagnosis, and recovery	3	1	0	1	0	1	1
Help and documentation	6	3	2	2	1	1	0
Total	45	23	8	10	8	11	8

<sup>a</sup>ΣA: total number of points.

<sup>b</sup>ΣB: number of points after removing duplicates.

**Figure 3.** Examples of pages from the MyShield self-care mobile app.



**MyShield Usability and Satisfaction**

Table 2 shows that the mean age of the 27 participants was 29.8 years; 12 (44%) participants were aged between 29 and 38 years. Fifteen participants (56%) were women; 19 (70%) had a bachelor’s degree or higher; and 12 (44%) had used a smartphone for at least 5 years.

The participants completed the MAUQ questionnaire designed for standalone mHealth apps. Overall, all items received a score of 3.5 or higher (Table 3). The “interface and satisfaction” component was rated the strongest with 95.9%, followed by “ease of use” (89.5%) and “usefulness” (86.7%).

**Table 2.** Demographic characteristics of the participants (N=27).

Characteristics, n (%)	Values
<b>Age (years)</b>	
18-28	7 (26)
29-38	12 (44)
39-50	8 (30)
<b>Gender</b>	
Male	12 (44)
Female	15 (56)
<b>Education</b>	
High school diploma	3 (11)
Some college credits, no degree	5 (19)
Bachelor's degree	10 (37)
Master's degree	5 (19)
Professional degree	3 (11)
Doctoral degree	1 (4)
<b>Duration of smartphone use (years)</b>	
1-2	5 (19)
3-5	10 (37)
5 or more	12 (44)

**Table 3.** The mean and SD of the results obtained from the MyShield usability app according to the Mobile Health App Usability Questionnaire. These statements can receive responses ranging from 1 (strongly disagree) to 7 (strongly agree).

Items, mean (SD) score	Values
<b>Ease of use</b>	
1. The app was easy to use.	6.54 (0.499)
2. It was easy for me to learn to use the app.	6.33 (0.471)
3. The interface of the app allowed me to use all the functions (such as entering information, responding to reminders, viewing information) offered by the app.	6.29 (0.568)
4. The navigation was consistent when moving between screens.	6.15 (0.422)
5. Whenever I made a mistake using the app, I could recover easily and quickly.	6.00 (0.756)
<b>Interface and satisfaction</b>	
1. I like the interface of the app.	6.95 (0.213)
2. The information in the app was well organized, so I could easily find the information I needed.	6.93 (0.258)
3. The app adequately acknowledged and provided information to let me know the progress of my action.	6.89 (0.348)
4. Overall, I am satisfied with this app.	6.83 (0.373)
5. The amount of time involved in using this app has been fitting for me.	6.69 (0.462)
6. I would use this app again.	6.49 (0.523)
7. I feel comfortable using this app in social settings.	6.18 (0.383)
<b>Usefulness</b>	
1. The app improved my access to health care services.	6.51 (0.5)
2. The app would be useful for my health and well-being.	6.44 (0.585)
3. The app helped me manage my health effectively.	6.20 (0.72)
4. I could use the app even when the internet connection was poor or not available.	5.88 (0.391)
5. This app has all the functions and capabilities I expected it to have.	5.75 (0.652)
6. This mHealth app provided an acceptable way to receive health care services, such as accessing educational materials, tracking my own activities, and performing self-assessment.	5.63 (0.483)

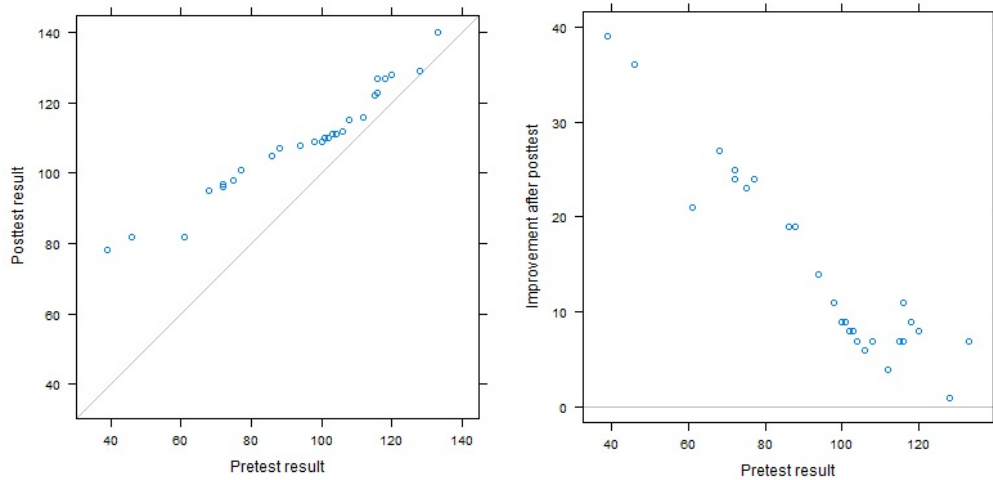
### Impact Rate of MyShield: Pretest and Posttest

In the pretest, the participants completed an SCI questionnaire before using the app: 2 respondents scored below 50, 8 scored 50 to 90, 14 scored 91 to 119, and 3 scored above 120 (Figure 4). In the posttest, the participants' self-care scores increased notably at low levels, but those at high self-care levels increased only a little: 2 respondents who had a score below 50 in the pretest stage reached scores above 50 (even reaching close to 90) (Figure 4).

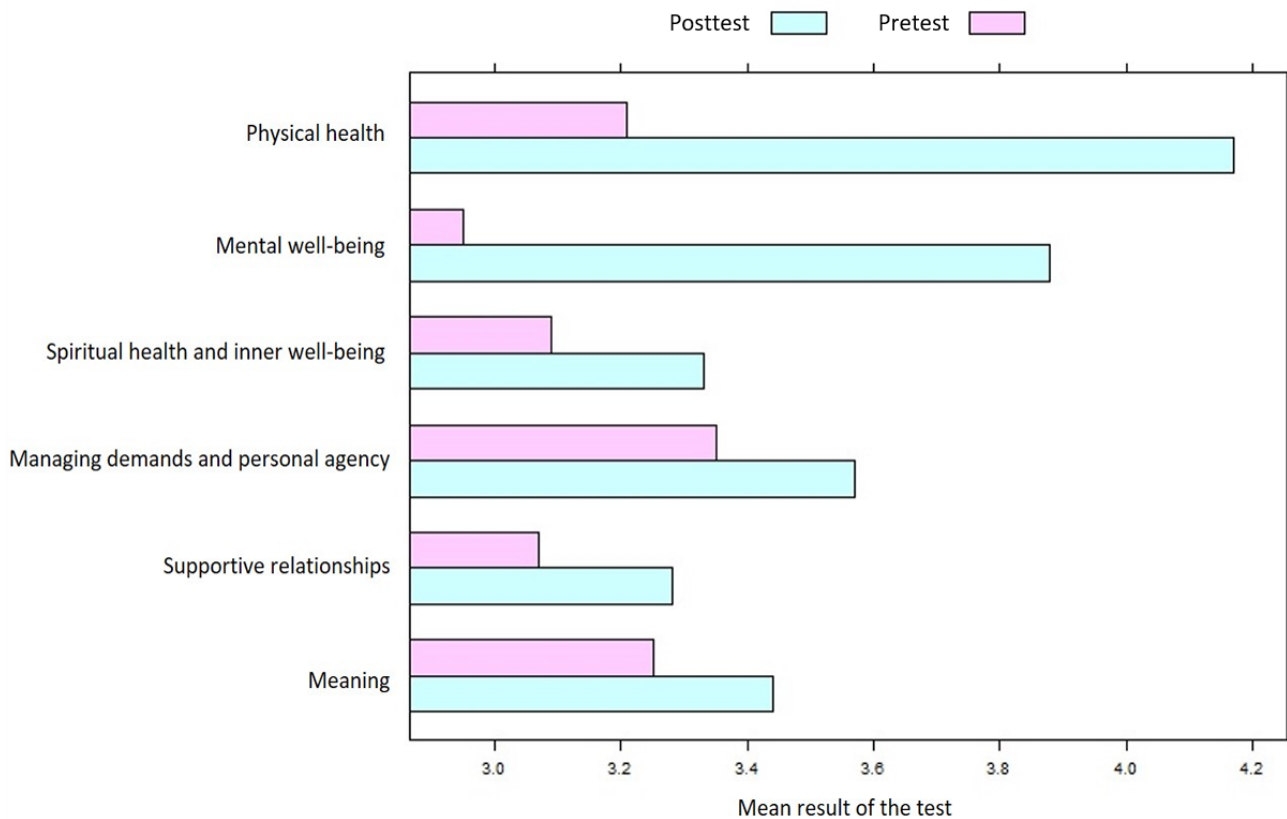
After the users used MyShield, we examined their average scores (0-5) for the self-care dimensions in the pretest and posttest to determine which dimensions had improved and which had decreased. The mean mental well-being score increased from 2.95 in the pretest to 3.88 in the posttest. For the physical health dimension, the mean score increased from 3.21 to 4.17. The increases for the other dimensions were visibly smaller (Figure 5).



**Figure 4.** Pre- and posttest participants' self-care assessment. The left panel shows posttest results against pretest results. Points above the gray line show an improvement in the posttest. The right panel (a modification of the modified Tukey mean-difference plot, presented by Kozak and Wnuk [32]) shows the improvement after the posttest against the pretest result.



**Figure 5.** Comparison of mean scores for the self-care dimensions of the MyShield app. The graph is ordered by the dimensions' improvement among the users.



## Discussion

### Principal Findings

This study aimed to develop a self-care app in order to help users improve their skills to prevent COVID-19 infection. To this end, we have developed, with the help of various field experts, a mobile app called MyShield; it is designed to work mainly in underdeveloped countries, as people in such countries are often left alone to struggle with their fear of COVID-19.

We designed MyShield using the ADDIE educational model. This is a novel, interdisciplinary approach to app design. MyShield is a mobile app that, in addition to being a potential public health resource, is also an educational entity built to teach people self-care skills and health knowledge that is effective well beyond the use of the app. MyShield utilizes the ADDIE educational model to enable users to teach themselves these self-care skills, making the app a channel of knowledge the user can learn from. A possible effect that the app attempts to achieve is an increased chance for users to protect themselves from COVID-19. While MyShield is just a tool to initiate this

learning, it is the source of the positive effects we have just discussed. We hope this plan will come to fruition, and that MyShield will encourage people from underdeveloped countries to believe in their own skills and learn or improve their self-care skills to reduce health risks.

As a standalone app, MyShield includes 6 major menus and various subcategories directly related to self-care training. In order to see how MyShield helped the participants, we conducted pretests before the initial use of the app. After participants used MyShield for 45 days, we conducted posttests to see if using the app helped them enhance their self-care skills. Both tests used a self-care inventory questionnaire to assess the participants' self-care levels. All participants showed improvement; those who had shown poor self-care skills in the pretest improved much more than those with higher self-care awareness (Figure 5).

### Comparison With Prior Work

In the COVID-19 era, mobile apps are particularly useful when their design does not limit their use to a particular group or particular members of society, but rather allows them to be used by all society's members, including young people, older adults, and people with physical disabilities. This issue is so crucial that it is emphasized as one of social justice and human rights [33,34]. To reach this goal, however, mobile apps need to be designed accordingly. In view of the user experience and user interface of MyShield, we tried to move the app user interface toward better usability and accessibility for various people, manipulating such features as font size and color, brightness, and dark and light themes [35]. We also tried to facilitate using the app so that the user did not need to perform complex functions. This led to an app design that did not require any special skills, such as working with Bluetooth or phone settings, and in which all functions were easily understood. We recognize one possible practical limitation of this study, which was the small number of participants, particularly older ones [36]. This suggests further work to be done in this area.

Islam et al [37] examined many mobile apps designed specifically for COVID-19. They classified the apps into various categories, such as remote assistance, patient monitoring, current status, COVID-19 prevention, COVID-19 control, communication support, and treatment services. None of these apps, however, directly emphasized training in self-care skills [37]. MyShield addresses this gap; it does so by directly focusing on the training of self-care skills. Thus, MyShield can be particularly useful in lower-income and underdeveloped countries, since studies have shown their citizens do not pay as much attention to training self-care skills [15]. This may result from many social and economic factors, even though this is perhaps the least expensive way to improve a society's condition, particularly in the case of COVID-19 [14,18]. The impact ratings of the MyShield app indicate that it can improve people's self-care skills by providing credible educational materials and enhancing citizens' motivation to maintain their physical health and personal hygiene during the COVID-19 era. According to our research, MyShield's ease of use rating reached 89.5%, a value suggesting that the app is easy to use. This includes easy learning, proper organization of information,

and convenient access to information, features that many studies have emphasized are of crucial importance. Indeed, as other researchers have shown, ease of use strongly affects users' perception of mHealth services [38-40]. Various studies have found that important items for assessing the quality of mHealth apps include information management, navigation consistency, guidelines, user interface design, error prevention, user control, freedom, performance speed, medical features, and content validity [41-44].

MyShield's strength is in its interface, which had a satisfaction rating of 95.9%. According to the participants, the app organized information well, displayed progress, provided health information in a comprehensive manner, and let its users perform all its functions in an appropriate way. Access to health care services and educational materials is a unique feature of health apps [18,45-47]. MyShield was also rated well in these areas by the participants, who reported that they had no problems with accessing educational materials, tracking their own activities, and performing self-assessment.

The MAUQ results showed that the app met the health and care needs of individuals and helped them to improve their self-care skills by providing them with relevant information. We consider that such a systematic approach, taking into account people's needs and limitations, reaches far beyond the confines of many contributions of COVID-19 self-care apps.

### Limitations

The study had a small number of participants. The number, however, was large enough to represent the app's user groups in aspects such as education, gender, and age. However, we could have specifically included more older participants, most of whom are unfamiliar with health care apps on smartphones. This would have required a much larger study group. Our study designed the app specifically to address the needs of underdeveloped countries, but we conducted our survey in Iran, a more developed country than the target audience of this app. However, most participants—selected through purposive sampling—had little knowledge about self-care, so they can be considered a fair representation of people from underdeveloped countries.

Another limitation was the length of time (45 days) during which the participants had the opportunity to use MyShield. Budget limitations prevented us from extending the study beyond 45 days. Perhaps if the participants used the MyShield app for a longer period of time, the self-care inventory would have improved. In addition, the app was available for Android only, automatically excluding all iOS (ie, Apple) users. A future study may address these issues.

### Conclusion

This study demonstrates that using a standalone mobile app based on a standard educational model can improve people's self-care skills. Improving self-care skills in underdeveloped countries with mobile health apps is a rather simple, inexpensive, and effective solution. It does, however, require a way to reach people and to convince them to use such an app. We consider that the design and development of MyShield can be used as a practical and realistic model for the development of mHealth

apps. According to our findings, standard educational models combined with existing designs (eg, user-centered or interactive designs) for mobile health apps can meet most requirements of mobile educational apps.

As our research shows, MyShield can help people from both underdeveloped and developing countries, where the fight

against COVID-19 is particularly difficult due to economics. We developed MyShield using a special educational design method to enable people to better learn to improve their self-care skills, thus helping to better protect themselves against COVID-19.

---

## Acknowledgments

The authors would like to thank all the professors and staff of the Tarbiat Modares University who helped us in this study. The second author (M Kozak)—like all the other authors—also thank Howard for his help with English-language editing of the text. In addition, the first author of this paper (HRS) as well as all other authors would like to thank Mr Afshin Babajani for his help with the design of the MyShield app. Finally, we consider it necessary to dedicate this research work to the medical staff and all those who have been risking their lives to save others. A grant from the Romanian National Authority for Scientific Research and Innovation to MA and CH was partially used to support this project (PN-III-P4-ID-PCCF-2016-0084).

---

## Authors' Contributions

HRS conceived and designed the mobile app and drafted the manuscript. ZM and AG collected and interpreted the data. M Kozak participated in designing the evaluation, performed parts of the statistical analysis, and helped revise the manuscript. MA and CH reevaluated the data, revised the manuscript, and performed the statistical analysis. M Karajizadeh and MH reanalyzed the statistical data. All authors read and approved the final manuscript.

---

## Conflicts of Interest

None declared.

---

Multimedia Appendix 1

Self-care inventory.

[[DOCX File, 20 KB - formative\\_v6i9e39718\\_app1.docx](#)]

---

## References

1. Zhao S, Musa S, Lin Q, Ran J, Yang G, Wang W, et al. Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak. *J Clin Med* 2020 Feb 01;9(2):388 [FREE Full text] [doi: [10.3390/jcm9020388](https://doi.org/10.3390/jcm9020388)] [Medline: [32024089](https://pubmed.ncbi.nlm.nih.gov/32024089/)]
2. Liu J, Zheng X, Tong Q, Li W, Wang B, Sutter K, et al. Overlapping and discrete aspects of the pathology and pathogenesis of the emerging human pathogenic coronaviruses SARS-CoV, MERS-CoV, and 2019-nCoV. *J Med Virol* 2020 May;92(5):491-494 [FREE Full text] [doi: [10.1002/jmv.25709](https://doi.org/10.1002/jmv.25709)] [Medline: [32056249](https://pubmed.ncbi.nlm.nih.gov/32056249/)]
3. Velavan T, Meyer C. The COVID-19 epidemic. *Trop Med Int Health* 2020 Mar;25(3):278-280 [FREE Full text] [doi: [10.1111/tmi.13383](https://doi.org/10.1111/tmi.13383)] [Medline: [32052514](https://pubmed.ncbi.nlm.nih.gov/32052514/)]
4. Ting D, Carin L, Dzau V, Wong T. Digital technology and COVID-19. *Nat Med* 2020 Apr;26(4):459-461 [FREE Full text] [doi: [10.1038/s41591-020-0824-5](https://doi.org/10.1038/s41591-020-0824-5)] [Medline: [32284618](https://pubmed.ncbi.nlm.nih.gov/32284618/)]
5. COVID Alert. Government of Canada. URL: <https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19/covid-alert.html> [accessed 2021-05-20]
6. COVIDSafe app. Commonwealth of Australia. URL: <https://www.health.gov.au/resources/apps-and-tools/covidsafe-app> [accessed 2022-09-07]
7. Saeidnia H, Ghorbi A, Kozak M, Herteliu C. Smartphone-based healthcare apps for older adults in the COVID-19 Era: heuristic evaluation. *Stud Health Technol Inform* 2022 Jan 14;289:128-131. [doi: [10.3233/SHTI210875](https://doi.org/10.3233/SHTI210875)] [Medline: [35062108](https://pubmed.ncbi.nlm.nih.gov/35062108/)]
8. Ali I, Ali S. Why may COVID-19 overwhelm low-income countries like Pakistan? *Disaster Med Public Health Prep* 2022 Feb;16(1):316-320 [FREE Full text] [doi: [10.1017/dmp.2020.329](https://doi.org/10.1017/dmp.2020.329)] [Medline: [32907694](https://pubmed.ncbi.nlm.nih.gov/32907694/)]
9. Dahab M, van Zandvoort K, Flasche S, Warsame A, Ratnayake R, Favas C, et al. COVID-19 control in low-income settings and displaced populations: what can realistically be done? *Confl Health* 2020;14:54 [FREE Full text] [doi: [10.1186/s13031-020-00296-8](https://doi.org/10.1186/s13031-020-00296-8)] [Medline: [32754225](https://pubmed.ncbi.nlm.nih.gov/32754225/)]
10. Josephson A, Kilic T, Michler J. Socioeconomic impacts of COVID-19 in low-income countries. *Nat Hum Behav* 2021 May;5(5):557-565. [doi: [10.1038/s41562-021-01096-7](https://doi.org/10.1038/s41562-021-01096-7)] [Medline: [33785897](https://pubmed.ncbi.nlm.nih.gov/33785897/)]
11. Brauer M, Zhao JT, Bennitt FB, Stanaway JD. Global access to handwashing: implications for COVID-19 control in low-income countries. *Environ Health Perspect* 2020 May;128(5):57005 [FREE Full text] [doi: [10.1289/EHP7200](https://doi.org/10.1289/EHP7200)] [Medline: [32438824](https://pubmed.ncbi.nlm.nih.gov/32438824/)]

12. Mahase E. Delta variant: What is happening with transmission, hospital admissions, and restrictions? *BMJ* 2021 Jun 15;373:n1513. [doi: [10.1136/bmj.n1513](https://doi.org/10.1136/bmj.n1513)] [Medline: [34130949](https://pubmed.ncbi.nlm.nih.gov/34130949/)]
13. Mollica R, Fernando D, Augusterfer E. Beyond burnout: responding to the COVID-19 pandemic challenges to self-care. *Curr Psychiatry Rep* 2021 Mar 09;23(4):21 [FREE Full text] [doi: [10.1007/s11920-021-01230-2](https://doi.org/10.1007/s11920-021-01230-2)] [Medline: [33728512](https://pubmed.ncbi.nlm.nih.gov/33728512/)]
14. Mehraeen E, Hayati B, Saeidi S, Heydari M, Seyedalinaghi S. Self-Care instructions for people not requiring hospitalization for coronavirus disease 2019 (COVID-19). *Arch Clin Infect Dis* 2020 Apr 02;15(COVID-19):1-6 [FREE Full text] [doi: [10.5812/archcid.102978](https://doi.org/10.5812/archcid.102978)]
15. Zareipour M, Kalejahi JN. The role of social participation in controlling and preventing of coronavirus 2019 disease in Iran. *Open Access Maced J Med Sci* 2020 Aug 15;8(T1):134-136 [FREE Full text] [doi: [10.3889/oamjms.2020.4956](https://doi.org/10.3889/oamjms.2020.4956)]
16. Fawcett J. The nurse theorists: 21st-century updates--Dorothea E. Orem. *Nurs Sci Q* 2001 Jan 18;14(1):34-38. [doi: [10.1177/08943180122108021](https://doi.org/10.1177/08943180122108021)] [Medline: [11873352](https://pubmed.ncbi.nlm.nih.gov/11873352/)]
17. Canady V. Amid COVID - 19, self - care as important for providers as it is for patients. *Ment Health Wkly* 2020 Apr 17;30(16):1-4 [FREE Full text] [doi: [10.1002/mhw.32324](https://doi.org/10.1002/mhw.32324)]
18. Saeidnia H, Mohammadzadeh Z, Saeidnia M, Mahmoodzadeh A, Ghorbani N, Hasanzadeh M. Identifying requirements of a self-care system on smartphones for preventing coronavirus disease 2019 (COVID-19). *Iran J Med Microbiol* 2020 May 01;14(3):241-251 [FREE Full text] [doi: [10.30699/ijmm.14.3.241](https://doi.org/10.30699/ijmm.14.3.241)]
19. Drew D, Nguyen L, Steves C, Menni C, Freydn M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: [10.1126/science.abc0473](https://doi.org/10.1126/science.abc0473)] [Medline: [32371477](https://pubmed.ncbi.nlm.nih.gov/32371477/)]
20. Kamel Boulos MN, Geraghty E. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *Int J Health Geogr* 2020 Mar 11;19(1):8 [FREE Full text] [doi: [10.1186/s12942-020-00202-8](https://doi.org/10.1186/s12942-020-00202-8)] [Medline: [32160889](https://pubmed.ncbi.nlm.nih.gov/32160889/)]
21. Novak LL, Unertl KM, Holden RJ. Realizing the potential of patient engagement: designing IT to support health in everyday life. *Stud Health Technol Inform* 2016;222:237-247 [FREE Full text] [doi: [10.3233/978-1-61499-635-4-237](https://doi.org/10.3233/978-1-61499-635-4-237)] [Medline: [27198106](https://pubmed.ncbi.nlm.nih.gov/27198106/)]
22. Branson R, Rayner G, Cox J, Furman J, King F, Florida State Univ Tallahassee Center For Educational Technology. Interservice procedures for instructional systems development: executive summary and model. Defense Technical Information Center. 1975. URL: <https://apps.dtic.mil/sti/pdfs/ADA019486.pdf> [accessed 2022-09-07]
23. Seels B, Richey R. *Instructional Technology: The Definition and Domains of the Field*. Charlotte, NC: Information Age Publishing; 2012.
24. Kurt S. ADDIE Model: Instructional Design. Educational Technology. URL: <https://educationaltechnology.net/the-addie-model-instructional-design/> [accessed 2022-09-07]
25. Saeidnia H, Ausloos M, Mohammadzadeh Z, Babajani A, Hassanzadhh M. Mobile-Based Self-Care Application for COVID-19: Development Process Using the ADDIE Model. *Stud Health Technol Inform* 2022 Jan 14;289:110-113. [doi: [10.3233/SHTI210871](https://doi.org/10.3233/SHTI210871)] [Medline: [35062104](https://pubmed.ncbi.nlm.nih.gov/35062104/)]
26. Murray P. Using Virtual Focus Groups in Qualitative Research. *Qual Health Res* 2016 Jul 01;7(4):542-549. [doi: [10.1177/104973239700700408](https://doi.org/10.1177/104973239700700408)]
27. Brown J, Nam Kim H. Usability evaluation of Alzheimer's mHealth applications for caregivers. *Proc Hum Factors Ergon Soc Annu Meet* 2018 Sep 25;62(1):503-507. [doi: [10.1177/1541931218621115](https://doi.org/10.1177/1541931218621115)]
28. Nielsen J. Enhancing the explanatory power of usability heuristics. In: CHI '94: SIGCHI Conference on Human Factors in Computing Systems. 1994 Apr 24 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 24, 1994; Boston, MA p. 152-158 URL: <https://dl.acm.org/doi/pdf/10.1145/191666.191729> [doi: [10.1145/191666.191729](https://doi.org/10.1145/191666.191729)]
29. Turney L, Pocknee C. Virtual focus groups: new frontiers in research. *Int J Qual Methods* 2016 Nov 29;4(2):32-43 [FREE Full text] [doi: [10.1177/160940690500400203](https://doi.org/10.1177/160940690500400203)]
30. Self-Care Inventory 2020. David Irvine: The Leader's Navigator. URL: <https://davidirvine.com/assessments/> [accessed 2022-09-07]
31. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. *JMIR Mhealth Uhealth* 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: [10.2196/11500](https://doi.org/10.2196/11500)] [Medline: [30973342](https://pubmed.ncbi.nlm.nih.gov/30973342/)]
32. Kozak M, Wnuk A. Including the Tukey mean-difference (Bland-Altman) plot in a statistics course. *Teach Stat* 2014 Jan 29;36(3):83-87. [doi: [10.1111/test.12032](https://doi.org/10.1111/test.12032)]
33. Walls D. User experience in social justice contexts. In: 34th ACM International Conference on the Design of Communication. 2016 Presented at: Proceedings of the 34th ACM International Conference on the Design of Communication; Sep 23, 2016; New York, NY p. 1-6 URL: <https://dl.acm.org/doi/abs/10.1145/2987592.2987604> [doi: [10.1145/2987592.2987604](https://doi.org/10.1145/2987592.2987604)]
34. Harihareswara S. User experience is a social justice issue. *Code4Lib Journal*. 2015. URL: <https://journal.code4lib.org/articles/10482> [accessed 2022-09-07]

35. Morris J. User interface design for older adults. *Interact Comput* 1994 Dec;6(4):373-393. [doi: [10.1016/0953-5438\(94\)90009-4](https://doi.org/10.1016/0953-5438(94)90009-4)]
36. Rizzo E. COVID-19 contact tracing apps: the 'elderly paradox'. *Public Health* 2020 Aug;185:127 [FREE Full text] [doi: [10.1016/j.puhe.2020.06.045](https://doi.org/10.1016/j.puhe.2020.06.045)] [Medline: [32622218](https://pubmed.ncbi.nlm.nih.gov/32622218/)]
37. Islam MN, Islam I, Munim KM, Islam AKMN. A review on the mobile applications developed for COVID-19: an exploratory analysis. *IEEE Access* 2020;8:145601-145610 [FREE Full text] [doi: [10.1109/ACCESS.2020.3015102](https://doi.org/10.1109/ACCESS.2020.3015102)] [Medline: [34812346](https://pubmed.ncbi.nlm.nih.gov/34812346/)]
38. Akter S, D'Ambra J, Ray P. User perceived service quality of mHealth services in developing countries. 2010 Presented at: 18th European Conference on Information Systems; Apr 4, 2010; Pretoria, South Africa p. 1-12 URL: <https://ro.uow.edu.au/commpapers/3128/>
39. Schnall R, Higgins T, Brown W, Carballo-Diequez A, Bakken S. Trust, Perceived Risk, Perceived Ease of Use and Perceived Usefulness as Factors Related to mHealth Technology Use. *Stud Health Technol Inform* 2015;216:467-471 [FREE Full text] [doi: [10.3233/978-1-61499-564-7-467](https://doi.org/10.3233/978-1-61499-564-7-467)] [Medline: [26262094](https://pubmed.ncbi.nlm.nih.gov/26262094/)]
40. Kayyali R, Peletidi A, Ismail M, Hashim Z, Bandeira P, Bonnah J. Awareness and use of mHealth apps: a study from england. *Pharmacy (Basel)* 2017 Jun 14;5(2):33 [FREE Full text] [doi: [10.3390/pharmacy5020033](https://doi.org/10.3390/pharmacy5020033)] [Medline: [28970445](https://pubmed.ncbi.nlm.nih.gov/28970445/)]
41. Yasini M, Beranger J, Desmarais P, Perez L, Marchand G. mHealth Quality: A Process to Seal the Qualified Mobile Health Apps. *Stud Health Technol Inform* 2016;228:205-209 [FREE Full text] [doi: [10.3233/978-1-61499-678-1-205](https://doi.org/10.3233/978-1-61499-678-1-205)] [Medline: [27577372](https://pubmed.ncbi.nlm.nih.gov/27577372/)]
42. Brown W, Yen P, Rojas M, Schnall R. Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. *J Biomed Inform* 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: [10.1016/j.jbi.2013.08.001](https://doi.org/10.1016/j.jbi.2013.08.001)] [Medline: [23973872](https://pubmed.ncbi.nlm.nih.gov/23973872/)]
43. 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](https://doi.org/10.2196/mhealth.5849)] [Medline: [27287964](https://pubmed.ncbi.nlm.nih.gov/27287964/)]
44. Saeidnia HR, Mohammadzadeh Z, Hassanzadeh M. Evaluation of mobile phone healthcare applications during the Covid-19 pandemic. *Stud Health Technol Inform* 2021 May 27;281:1100-1101. [doi: [10.3233/SHTI210363](https://doi.org/10.3233/SHTI210363)] [Medline: [34042856](https://pubmed.ncbi.nlm.nih.gov/34042856/)]
45. Gonçalves-Bradley DC, J Maria AR, Ricci-Cabello I, Villanueva G, Fønhus MS, Glenton C, et al. Mobile technologies to support healthcare provider to healthcare provider communication and management of care. *Cochrane Database Syst Rev* 2020 Aug 18;8:CD012927 [FREE Full text] [doi: [10.1002/14651858.CD012927.pub2](https://doi.org/10.1002/14651858.CD012927.pub2)] [Medline: [32813281](https://pubmed.ncbi.nlm.nih.gov/32813281/)]
46. VanHeuvelen JS. Isolation or interaction: healthcare provider experience of design change. *Sociol Health Illn* 2019 May;41(4):692-708. [doi: [10.1111/1467-9566.12850](https://doi.org/10.1111/1467-9566.12850)] [Medline: [30737809](https://pubmed.ncbi.nlm.nih.gov/30737809/)]
47. Global Burden of Disease 2020 Health Financing Collaborator Network. Tracking development assistance for health and for COVID-19: a review of development assistance, government, out-of-pocket, and other private spending on health for 204 countries and territories, 1990-2050. *Lancet* 2021 Oct 09;398(10308):1317-1343 [FREE Full text] [doi: [10.1016/S0140-6736\(21\)01258-7](https://doi.org/10.1016/S0140-6736(21)01258-7)] [Medline: [34562388](https://pubmed.ncbi.nlm.nih.gov/34562388/)]

## Abbreviations

**ADDIE:** analysis, design, development, implementation, and evaluation

**MAUQ:** Mobile Health App Usability Questionnaire

**mHealth:** mobile health

**MySQL:** My Structured Query Language

**SCI:** self-care inventory

*Edited by A Mavragani; submitted 19.05.22; peer-reviewed by S Subramaniam, Z Xie; comments to author 09.06.22; revised version received 13.07.22; accepted 10.08.22; published 13.09.22.*

*Please cite as:*

Saeidnia HR, Kozak M, Ausloos M, Herteliu C, Mohammadzadeh Z, Ghorbi A, Karajizadeh M, Hassanzadeh M

Development of a Mobile App for Self-Care Against COVID-19 Using the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) Model: Methodological Study

*JMIR Form Res* 2022;6(9):e39718

URL: <https://formative.jmir.org/2022/9/e39718>

doi: [10.2196/39718](https://doi.org/10.2196/39718)

PMID: [36054441](https://pubmed.ncbi.nlm.nih.gov/36054441/)

©Hamid Reza Saeidnia, Marcin Kozak, Marcel Ausloos, Claudiu Herteliu, Zahra Mohammadzadeh, Ali Ghorbi, Mehrdad Karajizadeh, Mohammad Hassanzadeh. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 13.09.2022.

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

Original Paper

# Personalized Energy Expenditure Estimation: Visual Sensing Approach With Deep Learning

Toby Perrett<sup>1</sup>, PhD; Alessandro Masullo<sup>1</sup>, PhD; Dima Damen<sup>1</sup>, PhD; Tilo Burghardt<sup>1</sup>, PhD; Ian Craddock<sup>1</sup>, PhD; Majid Mirmehdi<sup>1</sup>, PhD

University of Bristol, Bristol, United Kingdom

**Corresponding Author:**

Toby Perrett, PhD

University of Bristol

Digital Health, First Floor

1 Cathedral Square

Bristol, BS1 5DD

United Kingdom

Phone: 44 117 45 50375

Email: [toby.perrett@bristol.ac.uk](mailto:toby.perrett@bristol.ac.uk)

## Abstract

**Background:** Calorimetry is both expensive and obtrusive but provides the only way to accurately measure energy expenditure in daily living activities of any specific person, as different people can use different amounts of energy despite performing the same actions in the same manner. Deep learning video analysis techniques have traditionally required a lot of data to train; however, recent advances in few-shot learning, where only a few training examples are necessary, have made developing personalized models without a calorimeter a possibility.

**Objective:** The primary aim of this study is to determine which activities are most well suited to calibrate a vision-based personalized deep learning calorie estimation system for daily living activities.

**Methods:** The SPHERE (Sensor Platform for Healthcare in a Residential Environment) Calorie data set is used, which features 10 participants performing 11 daily living activities totaling 4.5 hours of footage. Calorimeter and video data are available for all recordings. A deep learning method is used to regress calorie predictions from video.

**Results:** Models are personalized with 32 seconds from all 11 actions in the data set, and mean square error (MSE) is taken against a calorimeter ground truth. The best single action for calibration is *wipe* (1.40 MSE). The best pair of actions are *sweep* and *sit* (1.09 MSE). This compares favorably to using a whole 30-minute sequence containing 11 actions to calibrate (1.06 MSE).

**Conclusions:** A vision-based deep learning energy expenditure estimation system for a wide range of daily living activities can be calibrated to a specific person with footage and calorimeter data from 32 seconds of sweeping and 32 seconds of sitting.

(JMIR Form Res 2022;6(9):e33606) doi:[10.2196/33606](https://doi.org/10.2196/33606)

## KEYWORDS

energy expenditure; calories, calorimetry; deep learning; computer vision

## Introduction

### Background

The ability to measure energy expenditure is important in a wide variety of settings. Examples range from sports training [1] to diabetes and cardiovascular disease monitoring [2]. Of particular interest is obesity management, where the amount of activity found in sedentary people at work and in the home can make a large difference to their overall fitness [3], especially when energy expenditure that is not due to exercise is taken into account [4]. The most accurate ways to measure

person-specific energy expenditure are to use a sealed chamber [5] or indirect calorimetry [6]. However, other than the upfront costs and time with such equipment, they are also intrusive and cumbersome when used for a significant length of time, and they require expert installation. Further, they are unsuitable for long-term deployment in homes (eg, for health monitoring applications), whether for large scale studies or for individual cases.

In the absence of such accurate measurements, clinicians have used metabolic equivalent task (MET) tables [7,8] as an approximation, where each action has an associated energy

expenditure value. This can be a time-consuming process, especially for a long sequence containing multiple activities, as each activity must be manually assigned start and end times. However, most importantly, METs are highly inaccurate compared to calorimetry. Hence, other approaches have sought to bridge the accuracy gap, while also reducing the burden on clinicians and annotators. For example, wearables have been explored as a cheaper, less intrusive, and more portable alternative [9-16] with improved results over METs. Large-scale home monitoring systems [17-19] have started to provide enough data to investigate computer vision approaches [20-22], which are cheap, much less intrusive, and more accurate. This provides the opportunity to extend the monitoring of energy consumption from stationary work environments [23-25], where variation between different people cannot be accurately captured by self-reporting.

The main problem with noncalorimeter-based approaches is that they still offer a *general model* only. That is, they will provide the same energy expenditure estimation for 2 individuals carrying out an action in a similar way, even though they may be using different amounts of energy.

Our aim is to estimate energy expenditure from observations of a person's physical movement. To this end, we train a deep learning model using footage of participants wearing calorimeters. Traditionally, deep learning methods have required a vast amount of data to personalize [26,27]. However, we exploit recent advances that can adapt general models to specific tasks [28-32] and determine which small set of actions is best suited to personalizing a general model. This will reduce the amount of calorimeter time per participant necessary for model personalization and will demonstrate that vision-based deep learning models are suitable for use in real-world settings. This is the first time in the literature a personalized vision-based energy expenditure estimation training regime has been addressed. On a more fundamental level, determining which actions are most suitable for fine-tuning a deep neural network can also give an indication about which types of activity are necessary to indicate a person's calorific profile. The approach introduced in this paper will be of practical use in many fields that monitor energy expenditure, such as sports training [1], nutrition [33], obesity management [34,35], and so on.

## Materials

For this study, we used the SPHERE (Sensor Platform for Healthcare in a Residential Environment) Calorie data set [36]. We briefly recap the key properties here before explaining our neural network approach to provide personalized energy expenditure estimations.

## Data Collection

A total of 10 participants performed a variety of daily living activities while using a K4b2 (COSMED) calorimeter. The

activities consisted of the following: *stand, sit, walk, wipe, vacuum, sweep, lie, exercise, stretch, clean, and read*. The participants are filmed using an off-the-shelf RGB-D (Red, Green, Blue plus Depth) sensor, and the video footage is pseudonymized by extracting silhouettes [37]. In total, 4.5 hours of footage at 30 frames per second and calorimeter data are available. To obtain a ground truth label for each video frame, calorimeter data are interpolated between each breath reading.

## Methods

### Ethics Approval

No ethics approval was required for this study, as we only used publicly available anonymized data for the purpose it was designed for.

### Overview

In this section, we will use our recently developed deep learning method [30] and provide a brief overview. Deep learning models consist of a neural network architecture, which processes a data stream (to give the energy expenditure estimation in our case) with an associated training regime to adapt a randomly initialized model to the desired task—often referred to as a learned model.

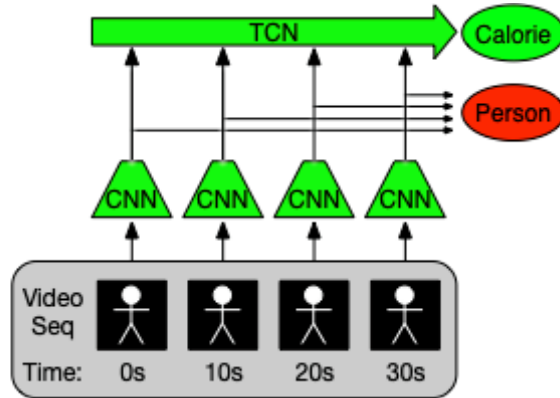
### Architecture

Deep neural network video architectures typically consist of 2 subnetworks, which are as follows: (1) a spatial subnetwork to extract useful features from each video frame—this part is necessary as the type of action currently being performed and the participant's body position can be an indicator of how much energy they are consuming. This is the convolutional neural network (Figure 1). Specifically, ResNet-18 [38] with pretrained ImageNet [39] weights is used; and (2) a temporal subnetwork to combine features extracted from each frame and to use this information to make an estimation—this part is necessary because just using 1 video frame is insufficient for energy consumption estimation; how fast participants move as well as their previous behavior and actions can have a great effect and must therefore be considered. For this stage, we deploy a temporal convolutional network [40] (Figure 1).

These 2 subnetworks are trained jointly (in this paper, we refer to this combined architecture as the “network”), so they can learn to specialize short- or long-term observations effectively. Previous works have shown that around 30 seconds of video footage is required to accurately regress calorie values [21,36] as previous activity affects the current calorimeter reading. Thus, we take advantage of an architecture that uses the spatial subnetwork to observe at 1 frame per second and a temporal subnetwork to combine 30 seconds worth of spatial subnetwork features.



**Figure 1.** Neural network architecture for processing silhouette video streams, consisting of a convolutional neural network (CNN) for extracting frame features and a temporal convolutional network (TCN) for combining frame features over a period of 30 seconds. To achieve an initialization that can be quickly adapted to unseen participants, the main training objective is to minimize the calorie loss while maximizing the person loss. Seq: sequence.



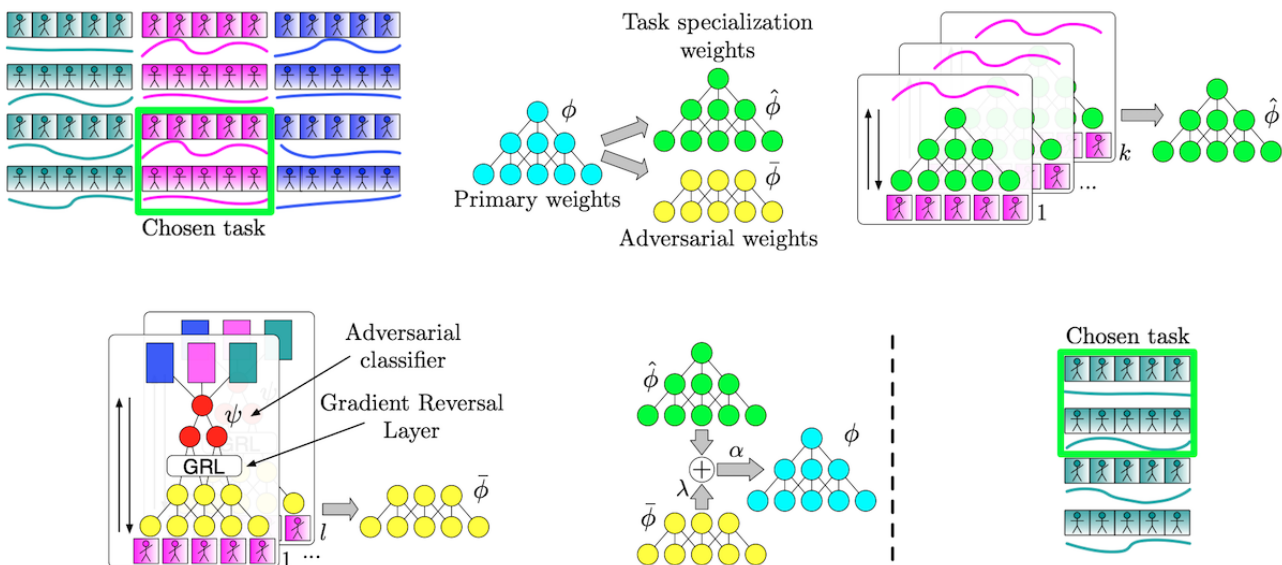
**Training**

Given an architecture to process the video data, along with silhouette videos and calorimeter readings, a training regime is required to learn from examples in a training set. As a large amount of data is usually required to train a neural network [26,27], they are often “pretrained” on a related large data set, then “fine-tuned” on the data set being used. However, in the case of learning a personalized model, the data requirements are still too large to be used for conventional fine-tuning. Thus, we use our recently developed few-shot (otherwise known as “meta-learning” or “learning to learn”) technique [30], which aims to learn a model that can be fine-tuned with very little amount of data.

Instead of optimizing the estimation of the current network, the training process optimizes the estimation of the network after it has been fine-tuned to a random participant from the training set, while an adversarial component aims to make the initialization agnostic to the participants in the training set. Figure 2 provides an illustration of this process. Specifically, it shows that first, a small “task” is constructed from the training set, containing a small amount of silhouette video and associated

calorimeter readings. Subsequently, 2 copies of the network initialization (ie, primary weights) are taken, which are named the task specialization and adversarial weights. The task specialization network is fine-tuned for a small number of iterations and becomes well suited to the current task. The adversarial weights are combined with an adversarial classifier, which are trained to predict which participant is used for this specific task. However, during this part, the gradients between the adversarial classifier and adversarial weights are negated. This means that as the adversarial classifier becomes better at classifying the person, the adversarial weights lose the ability to classify the person (ie, they become person agnostic). The task specialization and adversarial weights are finally merged back into the primary set of weights, and the process repeats with a different task. This process results in a set of primary weights that are agnostic to the participants in the training set yet are well suited to fine-tuning to unseen participants (Figure 2). For evaluation on an unseen participant, the primary set of weights are fine-tuned using a small amount of data from the unseen participant, and the adversarial component is not required because we want the evaluation network to be personalized to the evaluation participant.

**Figure 2.** Visualization of our data pipeline used to train and fine-tune a neural network, which is then used to provide personalized energy expenditure estimations from video.



## Results

### Overview

In this section, we outline our experiments and their results. Our aim is to find a network fine-tuning procedure that requires the minimum amount of data. In practice, this means less participant, clinician, and calorimeter time is required to personalize an energy expenditure model.

### Experimental Setup

A leave-one-out cross validation is used. In other words, 9 participants are used to train the model, and the 10th is used for evaluation. This process is repeated for each participant. To provide context to our results, we compare them against the following baselines: (1) MET value, which is calculated using expert labelled action start and end times; (2) no fine-tuning, which is a general model baseline as it can only make estimations with information learned from participants not being evaluated on; (3) fine-tune on one whole sequence of the participant the model is being personalized to—here, much more data are available to fine-tune than for the rest of our experiments, so this represents an upper bound for performance. The average length of a sequence is 30 minutes; (4) comparison with the work that introduces the meta-learning method in this paper [30], but only fine-tuned on the start of a sequence that contains frames without action labels; and (5) fine-tune using data from all 11 actions (32 seconds per action). This shows that standard training or fine-tuning fails with small amounts of data, even if fine-tuned with examples from all actions.

Mean square error (MSE) of the neural network estimation against the ground truth calorimeter reading is used as the evaluation metric. Note that an error is calculated for every video frame (but the model will have seen the previous 30 seconds of video to make this prediction).

There are 2 long (20-30 minutes) sequences per participant. For all experiments, the network is fine-tuned using data from sequence 1 and evaluated on sequence 2 and vice versa. This ensures that no data for evaluation have been seen during training or fine-tuning.

### Single-Action Personalization

To fine-tune to the participant being used for evaluation, 60 video clips are used. As we are assessing how well the model personalizes using a single action, these 60 clips are taken from a 32-second block of video where the fine-tuning action first appears. Each clip contains 30 uniformly sampled frames from 30 seconds of video (ie, sampling 1 frame every second). Given 32 seconds of video at 30 frames per second, there are  $32 \times 30 = 960$  frames. The first video clip uses frames 1, 31, ..., 901. The second video clip uses frames 2, 32, ..., 902, and so on.

The first row of [Table 1](#) shows the results of fine-tuning on each action compared against the baselines listed above. We can see that 32 seconds of *wipe* is best for learning a personalized calorific profile. However, it is still short of the upper bound on performance. The model fine-tuned on a whole video sequence has an MSE of 1.06 compared to 1.40 for *wipe*.

**Table 1.** Mean square error averaged across all participants.

Actions	Stand	Sit	Walk	Wipe	Vacuum	Sweep	Lie	Exercise	Stretch	Clean	Read
Single action	2.32	2.20	2.80	1.40 <sup>a</sup>	1.94	2.15	2.43	6.77	17.85	3.27	2.47
<b>Action pairs</b>											
Stand	2.55	— <sup>b</sup>	—	—	—	—	—	—	—	—	—
Sit	2.65	2.09	—	—	—	—	—	—	—	—	—
Walk	2.87	2.57	2.72	—	—	—	—	—	—	—	—
Wipe	1.52	1.50	1.72	1.55	—	—	—	—	—	—	—
Vacuum	1.40	1.77	1.74	1.34	2.01	—	—	—	—	—	—
Sweep	1.61	1.09 <sup>c</sup>	1.59	1.36	1.60	2.55	—	—	—	—	—
Lie	1.56	1.24	2.39	1.38	2.12	2.60	2.34	—	—	—	—
Exercise	2.42	1.87	2.82	3.18	2.89	2.62	3.50	5.72	—	—	—
Stretch	17.70	3.01	4.47	6.34	4.88	4.83	11.79	7.63	12.65	—	—
Clean	1.45	1.59	1.71	1.52	2.46	2.03	2.08	5.28	8.06	3.42	—
Read	1.98	4.98	2.47	1.40	2.16	2.24	2.44	3.43	3.02	2.42	2.35
<b>Baselines</b>											
MET <sup>d</sup>	2.87	—	—	—	—	—	—	—	—	—	—
Before train only	2.17	—	—	—	—	—	—	—	—	—	—
All actions (whole sequence)	1.06 <sup>e</sup>	—	—	—	—	—	—	—	—	—	—
All actions (32s/action)	3.30	—	—	—	—	—	—	—	—	—	—
Sequence start [30]	1.74	—	—	—	—	—	—	—	—	—	—

<sup>a</sup>Best single action.

<sup>b</sup>Not applicable.

<sup>c</sup>Best paired action.

<sup>d</sup>MET: metabolic equivalent task.

<sup>e</sup>Best baseline.

## Multiple Action Personalization

With the hypothesis that a broader range of actions provides a better calorific profiling of a person, we deploy multiple actions to fine-tune. This is motivated by the example in Figure 3 and the associated single-action personalization results, where fine-tuning on a whole sequence outperforms models fine-tuned on any single action. For the following experiments, we compare every pair of actions. For each action, the same amount of footage is available to fine-tune as there was in the previous experiments (ie, 32 seconds). Table 1 also shows the results of all 2-action combinations averaged per participant. To verify that any improvement is not just due to an increase in fine-tuning data (ie, 64 seconds from 2 actions compared to 32 seconds

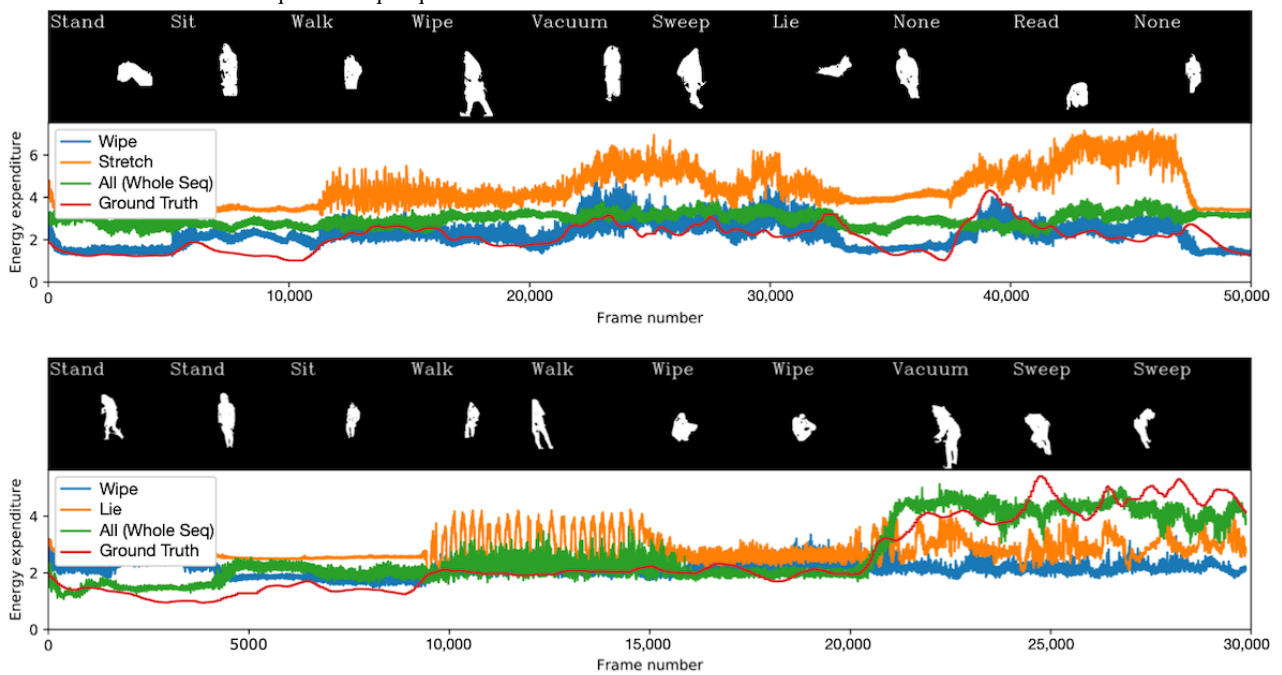
from 1), we include single-action results with the larger 64 seconds of fine-tuning data.

The best performing pair (*sweep* and *sit*) has an MSE of 1.09, which outperforms the best single-action pair (*wipe*, MSE 1.40). It is also very close to the whole sequence baseline, despite using much less data (64 seconds compared to 30 minutes).

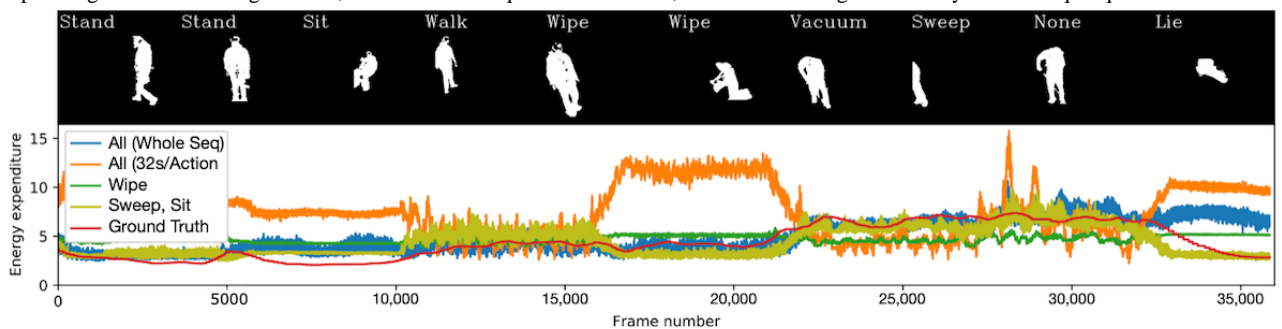
An example of multiple-action fine-tuning is given in Figure 4, for which the whole sequence model performs the worst. The best single (*wipe*) and pair (*sweep* and *sit*) fine-tuned models are shown alongside models fine-tuned on the whole sequence and with all actions (32 seconds per action).

Finally, Table 2 details the baselines, single-action results, and selected double-action results for each person individually.

**Figure 3.** Example energy expenditure estimations from silhouettes (recorded at 30 frames per second) using single action fine-tuning. The top example shows a success case where a model fine-tuned using only 32 seconds of wipe outperforms the whole sequence baseline, and that stretch is not a good action to use. The bottom example shows a failure case, where the models fine-tuned on a single action do not adapt to the period of high energy expenditure toward the end of a sequence. Seq: sequence.



**Figure 4.** An example sequence of silhouettes and energy expenditure estimations. Here, the best pair of actions for calibration across all participants is compared against the best single action, a whole video sequence to calibrate, and shorter footage from every action. Seq: sequence.



**Table 2.** Mean square error of baselines and single- and selected double-action fine-tuned models. The results are shown for each participant (“Pn”) individually along with the average over all participants. A blank entry indicates the action was not in video sequence used for fine-tuning.

Actions	Participants										Average
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
<b>Baselines</b>											
MET <sup>a</sup>	2.19	2.56	1.76	0.22	2.52	3.96	8.81	1.72	1.84	3.14	2.87
Before train only	1.38	0.82	0.87	0.69	1.46	3.55	7.41	0.71	1.43	3.34	2.17
All (whole sequence)	0.60	0.54	0.62	0.14	1.54	1.54	1.75	0.28	0.55	2.02	1.06 <sup>b</sup>
All (32s/action)	0.85	0.41	0.74	0.09	1.11	2.53	22.79	0.79	0.63	3.10	3.30
Sequence start [30]	0.29	0.58	0.54	0.29	1.25	2.30	3.24	3.50	0.65	4.73	1.74
<b>Single action</b>											
Stand	0.53	0.67	0.60	0.50	1.10	5.04	4.26	2.20	0.59	7.66	2.32
Sit	0.49	0.92	0.42	0.21	1.13	3.02	3.35	3.12	0.42	8.96	2.20
Walk	0.80	0.53	0.47	0.29	2.07	7.78	4.32	2.28	0.47	8.97	2.80
Wipe	0.29	1.36	0.45	0.36	0.73	3.37	2.95	1.80	0.48	2.17	1.40 <sup>c</sup>
Vacuum	0.79	0.63	0.54	0.60	1.67	2.95	5.18	1.89	0.85	4.29	1.94
Sweep	1.01	0.57	0.81	0.47	0.62	2.85	9.24	3.29	0.39	2.31	2.15
Lie	1.52	0.70	1.14	1.29	0.92	3.04	10.59	1.53	1.35	2.22	2.43
Exercise	1.56	0.76	— <sup>d</sup>	2.96	0.59	5.74	7.59	7.47	0.80	33.41	6.77
Stretch	5.93	46.52	0.48	5.82	5.16	21.19	30.64	13.86	30.81	18.05	17.85
Clean	1.17	2.15	0.94	0.32	1.04	5.93	8.94	2.17	4.65	5.35	3.27
Read	2.05	1.35	0.84	0.56	0.81	2.53	7.50	1.92	2.22	4.90	2.47
<b>Action pairs</b>											
Sweep/sit	0.96	0.67	0.47	0.13	0.90	2.51	1.02	0.99	0.47	2.75	1.09 <sup>e</sup>
Lie/sit	0.61	0.53	0.43	0.45	0.82	2.69	3.00	1.07	0.60	2.24	1.24
Vacuum/stand	0.87	0.57	0.38	0.14	1.53	3.39	1.73	0.78	0.61	4.04	1.40
Vacuum/wipe	0.48	0.64	0.64	0.19	1.21	3.25	1.59	1.63	0.60	3.15	1.34
Sweep/wipe	0.60	0.59	0.67	0.16	1.01	2.52	4.16	1.02	0.54	2.35	1.36
Wipe/wipe	0.57	0.95	0.48	0.11	1.07	3.88	3.43	1.98	0.56	2.42	1.55
Stretch/exercise	2.83	2.19	—	4.18	3.58	8.74	14.63	8.93	0.93	22.70	7.63
Clean/stretch	5.01	5.39	0.57	2.20	2.79	11.90	21.57	7.26	17.12	6.80	8.06
Stretch/lie	1.78	9.14	0.61	1.19	3.70	2.98	77.20	3.46	9.64	8.23	11.79
Stretch/stand	1.72	2.63	0.56	1.72	1.96	8.08	146.70	5.08	3.49	5.01	17.70

<sup>a</sup>MET: metabolic equivalent task.

<sup>b</sup>Best baseline.

<sup>c</sup>Best single action.

<sup>d</sup>Blank entries indicate the action was not in the video sequence used for fine-tuning.

<sup>e</sup>Best action pair.

## Discussion

### Single or Pair Difference

The results presented above raise several points for discussion. Perhaps the most important is why the best single action to fine-tune with (*wipe*) is not part of the best pair to fine-tune with (*sweep* and *sit*). Given a distribution of calorimeter or

silhouette sequences (which contain a wide variety of actions and calorific profiles), we would expect fine-tuning with 1 action to cover the middle of this distribution. If 2 actions are available, then each can be representative of more extreme parts of the energy expenditure or silhouette distribution while still adequately covering the middle of the distribution; 2 actions outperforming 1 corresponds to this intuition

## Action Variation

Another interesting observation is that there is a large amount of variation when fine-tuning using different actions. For example, fine-tuning using *stretch* is much worse than any other single action (17.85 MSE compared to the baseline 1.06 and second worst 6.77). One possible reason is that a participant stretching produces very different silhouettes compared to any of the other actions they perform. If a model is fine-tuned using these silhouettes, it has been conditioned to very different data compared to the other actions and thus gives bad estimations. A similar effect can be seen with *exercise*, which has less extreme but different silhouettes (6.77 MSE). This reasoning also applies to specific actions outperforming models fine-tuned on the sequence start. The sequence start may not provide enough information about a participant's calorific profile for the fine-tuned model to work well across a wide variety of actions.

## Participant Variation

There is also a difference in how all methods perform on specific participants. In particular, all models struggle on P10, with even the model fine-tuned on a whole sequence giving an MSE of 2.02. This is unlikely to be caused by visual differences (in the way that models fine-tuned on *stretch* are) as all actions perform poorly. Rather, it is most likely due to P10 having a calorific profile, which is very dissimilar to those found in all the other participants and could possibly be remedied by collecting data

from more participants to use during the training of the initialization.

## Conclusion

In this paper, we showed that a personalized calorie expenditure model that is more accurate than other existing techniques (bar intrusive calorimetry devices) is possible using a vision-based deep learning technique. The method can be personalized and can perform indefinitely in clinical and home environments after just 64 seconds of calorimeter calibration.

Our method uses a state-of-the-art deep learning technique, which learns an initialization from a data set containing calorimeter readings of footage from multiple participants. The initialization can then be adapted quickly to a participant unseen in the training set with footage and calorimeter readings of them *sweeping* for 32 seconds and *sitting* for 32 seconds. This personalized model outperforms the general models that have been used in the past.

The method outlined in this paper provides some benefits. It is suitable for long-term continuous monitoring of energy expenditure in daily-living scenarios and environments as it is noninvasive and does not require any change to participant behavior. It requires very little expensive clinician and calorimeter time to personalize, and it only needs a relatively cheap RGB-D sensor. Further, it does not require any human annotation of actions or activities after recording has finished.

## Acknowledgments

This work was performed under the SPHERE (Sensor Platform for Healthcare in a Residential Environment) Next Steps Project funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/R005273/1.

## Conflicts of Interest

None declared.

## References

1. Vogt S, Heinrich L, Schumacher YO, Grosshauser M, Blum A, König D, et al. Energy intake and energy expenditure of elite cyclists during pre-season training. *Int J Sports Med* 2005 Oct;26(8):701-706. [doi: [10.1055/s-2004-830438](https://doi.org/10.1055/s-2004-830438)] [Medline: [16158378](https://pubmed.ncbi.nlm.nih.gov/16158378/)]
2. Hamilton MT, Hamilton DG, Zderic TW. Role of low energy expenditure and sitting in obesity, metabolic syndrome, type 2 diabetes, and cardiovascular disease. *Diabetes* 2007 Nov;56(11):2655-2667 [FREE Full text] [doi: [10.2337/db07-0882](https://doi.org/10.2337/db07-0882)] [Medline: [17827399](https://pubmed.ncbi.nlm.nih.gov/17827399/)]
3. Epstein LH, Paluch RA, Consalvi A, Riordan K, Scholl T. Effects of manipulating sedentary behavior on physical activity and food intake. *J Pediatr* 2002 Mar;140(3):334-339. [doi: [10.1067/mpd.2002.122395](https://doi.org/10.1067/mpd.2002.122395)] [Medline: [11953732](https://pubmed.ncbi.nlm.nih.gov/11953732/)]
4. Villablanca PA, Alegria JR, Mookadam F, Holmes DR, Wright RS, Levine JA. Nonexercise activity thermogenesis in obesity management. *Mayo Clin Proc* 2015 Apr;90(4):509-519. [doi: [10.1016/j.mayocp.2015.02.001](https://doi.org/10.1016/j.mayocp.2015.02.001)] [Medline: [25841254](https://pubmed.ncbi.nlm.nih.gov/25841254/)]
5. Ravussin E, Lillioja S, Anderson TE, Christin L, Bogardus C. Determinants of 24-hour energy expenditure in man. Methods and results using a respiratory chamber. *J. Clin. Invest* 1986 Dec 1;78(6):1568-1578. [doi: [10.1172/jci112749](https://doi.org/10.1172/jci112749)]
6. Ferrannini E. The theoretical bases of indirect calorimetry: A review. *Metabolism* 1988 Mar;37(3):287-301. [doi: [10.1016/0026-0495\(88\)90110-2](https://doi.org/10.1016/0026-0495(88)90110-2)]
7. Byrne NM, Hills AP, Hunter GR, Weinsier RL, Schutz Y. Metabolic equivalent: one size does not fit all. *J Appl Physiol* (1985) 2005 Sep;99(3):1112-1119 [FREE Full text] [doi: [10.1152/jappphysiol.00023.2004](https://doi.org/10.1152/jappphysiol.00023.2004)] [Medline: [15831804](https://pubmed.ncbi.nlm.nih.gov/15831804/)]
8. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000 Sep;32(9 Suppl):S498-S504. [doi: [10.1097/00005768-200009001-00009](https://doi.org/10.1097/00005768-200009001-00009)] [Medline: [10993420](https://pubmed.ncbi.nlm.nih.gov/10993420/)]

9. Berntsen S, Hageberg R, Aandstad A, Mowinckel P, Anderssen SA, Carlsen K, et al. Validity of physical activity monitors in adults participating in free-living activities. *Br J Sports Med* 2010 Jul 15;44(9):657-664. [doi: [10.1136/bjism.2008.048868](https://doi.org/10.1136/bjism.2008.048868)] [Medline: [18628358](https://pubmed.ncbi.nlm.nih.gov/18628358/)]
10. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, et al. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. *Br J Sports Med* 2020 Mar 07;54(6):332-340. [doi: [10.1136/bjsports-2018-099643](https://doi.org/10.1136/bjsports-2018-099643)] [Medline: [30194221](https://pubmed.ncbi.nlm.nih.gov/30194221/)]
11. Conger SA, Scott SN, Bassett DR. Predicting energy expenditure through hand rim propulsion power output in individuals who use wheelchairs. *Br J Sports Med* 2014 Jul 13;48(13):1048-1053. [doi: [10.1136/bjsports-2014-093540](https://doi.org/10.1136/bjsports-2014-093540)] [Medline: [24825852](https://pubmed.ncbi.nlm.nih.gov/24825852/)]
12. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Med Sci Sports Exerc* 2016 Mar;48(3):457-465 [FREE Full text] [doi: [10.1249/MSS.0000000000000778](https://doi.org/10.1249/MSS.0000000000000778)] [Medline: [26484953](https://pubmed.ncbi.nlm.nih.gov/26484953/)]
13. Kuffel EE, Crouter SE, Haas JD, Frongillo EA, Bassett DR. Validity of estimating minute-by-minute energy expenditure of continuous walking bouts by accelerometry. *Int J Behav Nutr Phys Act* 2011 Aug 24;8(1):92 [FREE Full text] [doi: [10.1186/1479-5868-8-92](https://doi.org/10.1186/1479-5868-8-92)] [Medline: [21864359](https://pubmed.ncbi.nlm.nih.gov/21864359/)]
14. O'Driscoll R, Turicchi J, Hopkins M, Horgan GW, Finlayson G, Stubbs JR. Improving energy expenditure estimates from wearable devices: A machine learning approach. *J Sports Sci* 2020 Jul 06;38(13):1496-1505. [doi: [10.1080/02640414.2020.1746088](https://doi.org/10.1080/02640414.2020.1746088)] [Medline: [32252598](https://pubmed.ncbi.nlm.nih.gov/32252598/)]
15. O'Driscoll R, Turicchi J, Hopkins M, Duarte C, Horgan GW, Finlayson G, et al. Comparison of the validity and generalizability of machine learning algorithms for the prediction of energy expenditure: validation study. *JMIR Mhealth Uhealth* 2021 Aug 04;9(8):e23938 [FREE Full text] [doi: [10.2196/23938](https://doi.org/10.2196/23938)] [Medline: [34346890](https://pubmed.ncbi.nlm.nih.gov/34346890/)]
16. Gilgen-Ammann R, Schweizer T, Wyss T. Accuracy of the multisensory wristwatch polar vantage's estimation of energy expenditure in various activities: instrument validation study. *JMIR Mhealth Uhealth* 2019 Oct 02;7(10):e14534 [FREE Full text] [doi: [10.2196/14534](https://doi.org/10.2196/14534)] [Medline: [31579020](https://pubmed.ncbi.nlm.nih.gov/31579020/)]
17. Elsts A, Burghardt T, Byrne D, Camplani M, Damen D, Fafoutis X, et al. A guide to the SPHERE 100 homes study dataset. arXiv. 2018. URL: <http://arxiv.org/abs/1805.11907> [accessed 2022-07-05]
18. Das S, Dai R, Koperski M, Minciullo L, Garattoni L, Bremond F, et al. Toyota Smarthome: Real-World Activities of Daily Living. 2019 Presented at: International Conference on Computer Vision (ICCV); October 27, 2019 - November 02, 2019; Seoul, South Korea. [doi: [10.1109/ICCV.2019.00092](https://doi.org/10.1109/ICCV.2019.00092)]
19. Salatino C, Gower V, Ghrissi M, Tapus A, Wieczorowska-Tobis K, Suwalska A, et al. The EnrichMe Project. In: *Computers Helping People with Special Needs. ICCHP 2016. Lecture Notes in Computer Science*. Cham, Switzerland: Springer; 2016:326-334.
20. Tao L, Burghardt T, Mirmehdi M, Damen D, Cooper A, Camplani M, et al. Energy expenditure estimation using visual and inertial sensors. *IET Computer Vision* 2017 Oct 27;12(1):36-47. [doi: [10.1049/iet-cvi.2017.0112](https://doi.org/10.1049/iet-cvi.2017.0112)]
21. Masullo T, Burghardt T, Damen D, Hannuna S, Ponce-López V, Mirmehdi M. CaloriNet: From Silhouettes to Calorie Estimation in Private Environments. arXiv. 2018. URL: <https://arxiv.org/abs/1806.08152> [accessed 2022-07-05]
22. Nakamura S, Yeung S, Alahi A, Fei-Fei L. Jointly Learning Energy Expenditures and Activities Using Egocentric Multimodal Signals. 2017 Presented at: Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI, USA p. 6817-6826. [doi: [10.1109/cvpr.2017.721](https://doi.org/10.1109/cvpr.2017.721)]
23. Levine JA, Miller JM. The energy expenditure of using a "walk-and-work" desk for office workers with obesity. *Br J Sports Med* 2007 Sep 01;41(9):558-561 [FREE Full text] [doi: [10.1136/bjism.2006.032755](https://doi.org/10.1136/bjism.2006.032755)] [Medline: [17504789](https://pubmed.ncbi.nlm.nih.gov/17504789/)]
24. McAlpine DA, Manohar CU, McCrady SK, Hensrud D, Levine JA. An office-place stepping device to promote workplace physical activity. *Br J Sports Med* 2007 Dec 18;41(12):903-907 [FREE Full text] [doi: [10.1136/bjism.2006.034900](https://doi.org/10.1136/bjism.2006.034900)] [Medline: [17513333](https://pubmed.ncbi.nlm.nih.gov/17513333/)]
25. Kim SH, Kim JW, Park J, Shin MJ, Choi M. Predicting energy expenditure during gradient walking with a foot monitoring device: model-based approach. *JMIR Mhealth Uhealth* 2019 Oct 23;7(10):e12335 [FREE Full text] [doi: [10.2196/12335](https://doi.org/10.2196/12335)] [Medline: [31647467](https://pubmed.ncbi.nlm.nih.gov/31647467/)]
26. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 2017 May 24;60(6):84-90. [doi: [10.1145/3065386](https://doi.org/10.1145/3065386)]
27. Carreira J, Zisserman A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. 2017 Presented at: Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI, USA p. 4724-4733. [doi: [10.1109/cvpr.2017.502](https://doi.org/10.1109/cvpr.2017.502)]
28. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. arXiv. 2017. URL: <https://arxiv.org/abs/1703.03400?context=cs> [accessed 2022-07-05]
29. Snell K, Swersky K, Zemel R. Prototypical Networks for Few-Shot Learning. University of Toronto. 2017. URL: [https://www.cs.toronto.edu/~zemel/documents/prototypical\\_networks\\_nips\\_2017.pdf](https://www.cs.toronto.edu/~zemel/documents/prototypical_networks_nips_2017.pdf) [accessed 2022-07-05]
30. Perret T, Masullo A, Burghardt T, Mirmehdi M, Damen D. Meta-learning with context-agnostic initialisations. In: *Asian Conference on Computer Vision*. Cham, Switzerland: Springer; 2021:70-86.

31. Cao K, Ji J, Cao Z, Chang C, Niebles JC. Few-shot video classification via temporal alignment. Conference on Computer Vision and Pattern Recognition (CVPR) 2022:10615-10624. [doi: [10.1109/cvpr42600.2020.01063](https://doi.org/10.1109/cvpr42600.2020.01063)]
32. Perrett T, Masullo A, Burghardt T, Mirmehdi M, Damen D. Temporal-Relational CrossTransformers for Few-Shot Action Recognition. arXiv. 2021. URL: <http://arxiv.org/abs/2101.06184> [accessed 2022-07-05]
33. Hopkins M, Blundell JE, King NA. Individual variability in compensatory eating following acute exercise in overweight and obese women. Br J Sports Med 2014 Oct 10;48(20):1472-1476. [doi: [10.1136/bjsports-2012-091721](https://doi.org/10.1136/bjsports-2012-091721)] [Medline: [23666018](https://pubmed.ncbi.nlm.nih.gov/23666018/)]
34. Ehsam R, Hoerler-Koerner U, Stoffel S, Melges T, Ainsworth B. Exercise prescription for the overweight and the obese: how to quantify and yet keep it simple. Br J Sports Med 2009 Dec 16;43(12):951-953. [doi: [10.1136/bjism.2008.055657](https://doi.org/10.1136/bjism.2008.055657)] [Medline: [19293164](https://pubmed.ncbi.nlm.nih.gov/19293164/)]
35. Woo R, Garrow JS, Pi-Sunyer FX. Effect of exercise on spontaneous calorie intake in obesity. Am J Clin Nutr 1982 Sep;36(3):470-477. [doi: [10.1093/ajcn/36.3.470](https://doi.org/10.1093/ajcn/36.3.470)] [Medline: [7113952](https://pubmed.ncbi.nlm.nih.gov/7113952/)]
36. Tao L, Burghardt T, Mirmehdi M, Damen D, Cooper A, Hannuna S, et al. Calorie counter: RGB-depth visual estimation of energy expenditure at home. In: Computer Vision – ACCV 2016 Workshops. Cham, Switzerland: Springer; 2016:239-251.
37. Cao Z, Hidalgo G, Simon T, Wei S, Sheikh Y. OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell 2021 Jan 1;43(1):172-186. [doi: [10.1109/tpami.2019.2929257](https://doi.org/10.1109/tpami.2019.2929257)]
38. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 Presented at: Conference on Computer Vision and Pattern Recognition (CVPR); December 12, 2016; Las Vegas, NV, USA. [doi: [10.1109/cvpr.2016.90](https://doi.org/10.1109/cvpr.2016.90)]
39. Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. 2009 Presented at: Conference on Computer Vision and Pattern Recognition; August 18, 2009; Miami, FL, USA. [doi: [10.1109/cvpr.2009.5206848](https://doi.org/10.1109/cvpr.2009.5206848)]
40. Bai S, Kolter JZ, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv. 2018. URL: <https://arxiv.org/abs/1803.01271> [accessed 2022-07-05]

## Abbreviations

**MET:** metabolic equivalent task

**MSE:** mean square error

**RGB-D:** Red, Green, Blue plus Depth

**SPHERE:** Sensor Platform for Healthcare in a Residential Environment

*Edited by A Mavragani; submitted 15.09.21; peer-reviewed by G Liu, G Lim; comments to author 19.01.22; revised version received 15.03.22; accepted 25.03.22; published 14.09.22.*

*Please cite as:*

*Perrett T, Masullo A, Damen D, Burghardt T, Craddock I, Mirmehdi M*

*Personalized Energy Expenditure Estimation: Visual Sensing Approach With Deep Learning*

*JMIR Form Res 2022;6(9):e33606*

*URL: <https://formative.jmir.org/2022/9/e33606>*

*doi: [10.2196/33606](https://doi.org/10.2196/33606)*

*PMID: [36103223](https://pubmed.ncbi.nlm.nih.gov/36103223/)*

©Toby Perrett, Alessandro Masullo, Dima Damen, Tilo Burghardt, Ian Craddock, Majid Mirmehdi. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 14.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Digital Global Recruitment for Women's Health Research: Cross-sectional Study

Erika Rodriguez<sup>1\*</sup>, MS; Komal Peer<sup>1\*</sup>, PhD; Victoria Fruh<sup>1</sup>, PhD; Kaitlyn James<sup>2</sup>, MPH, PhD; Anna Williams<sup>1</sup>, MS; Alexis de Figueiredo Veiga<sup>1</sup>, MS; Michael R Winter<sup>3</sup>, MPH; Amanda Shea<sup>4</sup>, PhD; Ann Aschengrau<sup>5</sup>, SCD; Kevin J Lane<sup>6</sup>, PhD; Shruthi Mahalingaiah<sup>1,5,7</sup>, MD, MS

<sup>1</sup>Department of Environmental Health, Harvard TH Chan School of Public Health, Boston, MA, United States

<sup>2</sup>Deborah Kelly Center for Outcomes Research, Department of Obstetrics and Gynecology, Massachusetts General Hospital, Boston, MA, United States

<sup>3</sup>Biostatistics and Epidemiology Data Analytics Center, Boston University School of Public Health, Boston, MA, United States

<sup>4</sup>Clue by BioWink GmbH, Berlin, Germany

<sup>5</sup>Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

<sup>6</sup>Department of Environmental Health, Boston University School of Public Health, Boston, MA, United States

<sup>7</sup>Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, Massachusetts General Hospital, Boston, MA, United States

\*these authors contributed equally

**Corresponding Author:**

Shruthi Mahalingaiah, MD, MS

Department of Environmental Health

Harvard TH Chan School of Public Health

665 Huntington Avenue

Boston, MA, 02115

United States

Phone: 1 6174147305

Email: [shruthi@hsph.harvard.edu](mailto:shruthi@hsph.harvard.edu)

## Abstract

**Background:** With the increased popularity of mobile menstrual tracking apps and boosted Facebook posts, there is a unique opportunity to recruit research study participants from across the globe via these modalities to evaluate women's health. However, no studies to date have assessed the feasibility of using these recruitment sources for epidemiological research on ovulation and menstruation.

**Objective:** The objective of this study was to assess the feasibility of recruiting a diverse sample of women to an epidemiological study of ovulation and menstruation (OM) health (OM Global Health Study) using digital recruitment sources. The feasibility and diversity were assessed via click and participation rates, geographic location, BMI, smoking status, and other demographic information.

**Methods:** Participants were actively recruited via in-app messages using the menstrual tracking app Clue (BioWink GmbH) and a boosted Facebook post by DivaCup (Diva International Inc.). Other passive recruitment methods also took place throughout the recruitment period (eg, email communications, blogs, other social media). The proportion of participants who visited the study website after viewing and clicking the hypertext link (click rates) in the in-app messages and boosted Facebook post and the proportion of participants who completed the surveys per the number of completed consent and eligibility screeners (participation rates) were used to quantify the success of recruiting participants to the study website and study survey completion, respectively. Survey completion was defined as finishing the pregnancy and birth history section of the OM Global Health Study questionnaire.

**Results:** The recruitment period was from February 27, 2018, through January 24, 2020. In-app messages and the boosted Facebook post were seen by 104,000 and 21,400 people, respectively. Overall, 215 participants started the OM Global Health Study survey, of which 140 (65.1%), 39 (18.1%), and 36 (16.8%) participants were recruited via the app, the boosted Facebook post, and other passive recruitment methods, respectively. The click rate via the app was 18.9% (19,700 clicks/104,000 ad views) and 1.6% via the boosted Facebook post (340 clicks/21,400 ad views.) The overall participation rate was 44.6% (198/444), and the average participant age was 21.8 (SD 6.1) years. In terms of geographic and racial/ethnic diversity, 91 (44.2%) of the participants resided outside the United States and 147 (70.7%) identified as non-Hispanic White. In-app recruitment produced the most

geographically diverse stream, with 44 (32.8%) of the 134 participants in Europe, 77 (57.5%) in North America, and 13 (9.8%) in other parts of the world. Both human error and nonhuman procedural breakdowns occurred during the recruitment process, including a computer programming error related to age eligibility and a hacking attempt by an internet bot.

**Conclusions:** In-app messages using the menstrual tracking app Clue were the most successful method for recruiting participants from many geographic regions and producing the greatest numbers of started and completed surveys. This study demonstrates the utility of digital recruitment to enroll participants from diverse geographic locations and provides some lessons to avoid technical recruitment errors in future digital recruitment strategies for epidemiological research.

(*JMIR Form Res* 2022;6(9):e39046) doi:[10.2196/39046](https://doi.org/10.2196/39046)

## KEYWORDS

digital recruitment; internet; menstrual tracking app; menstrual; menstruation; reproductive health; reproduction; mobile health; menstrual health; mHealth; women's health; Facebook; social media; epidemiology research; in-app message; tracking app; health application; health app; eHealth; digital health; health technology; ovulation; recruit; attrition; research subject; participation; participant

## Introduction

Increased access to the internet via smartphones allows individuals to obtain information to better understand their menstrual cycles via social media, content-specific blogs, and mobile health apps. As of 2021, 93% of individuals in the United States use the internet. This usage is consistent across racial/ethnic groups, with 91% of Black Americans, 93% of White Americans, and 95% of Hispanic Americans using the internet [1]. In January 2018, 95% of Americans owned a mobile phone and 77% of those owned a smartphone. These numbers have steadily increased. In February 2021, 97% of Americans owned a mobile phone and 85% owned a smartphone [2]. Not only has access to these devices increased, but as of 2020, there were more than 300 reproductive health apps in the Apple App and Google Play Stores largely targeting women of childbearing age [3]. In this paper, we refer to individuals who menstruate as “women,” but we acknowledge that not all of those who menstruate identify as women. In 2019, the 5 most popular mobile menstrual tracking apps, according to obstetricians/gynecologists interviewed for an article in *Women's Health*, were Clue (BioWink GmbH), Flo (Flo Health Inc.), Ovia Fertility (Ovia Health), Eve by Glow (Glow Inc), and MagicGirl [4]. Recruitment via mobile menstrual tracking apps presents unique opportunities to advance epidemiological research on menstruation [5,6].

Researchers are beginning to take advantage of health apps and online platforms, including blogs, for recruitment of study participants [7]. In particular, the use of boosted Facebook posts for recruitment to health research is increasing in popularity, and recent studies have examined their use to recruit for clinical trials [8,9] and hard-to-reach populations [10]. Hard-to-reach populations include those who are traditionally underrepresented in research studies, such as people from racial/ethnic minorities [11] and rural populations [12]. The feminine hygiene market has an estimated worth of US \$35.4-\$40 billion [13] and presents an exciting opportunity to recruit a large, diverse, and global population of individuals who menstruate in order to better understand the factors that may impact the hypothalamic-pituitary-ovarian axis. Similarly, the women's health app market is estimated to be worth US \$20.8 billion [14]. However, studies assessing the utility of these sources for

recruitment to epidemiological research on ovulation and menstruation (OM) are lacking.

The ovulation and menstruation study (OM Global Health Study) recruited individuals from the mobile menstrual tracking app Clue and a boosted Facebook post by DivaCup (Diva International Inc.). Clue is a menstrual tracking app that was founded in 2013 by the Berlin-based company BioWink GmbH. It has over 12 million users from 190 different countries [15] and is considered to have 1 of the largest user bases worldwide among reproductive health apps [3]. Founded in 2001, Diva International Inc. is a Canadian menstrual care company that sells its menstrual cups in over 35 countries [16] and maintains engagement with its consumer base via its social media accounts, blog, and newsletter communications.

Previous research, such as that by Fenner et al [17], Wise et al [18], and Mahalingaiah et al [19], are examples of studies that recruited women through various digital modalities. Fenner et al [17] ran Facebook advertisements over a 5-month period targeting women in Victoria, Australia [17]. Wise et al [18] described recruitment for the Pregnancy Study Online (PRESTO) study over a 99-week period via internet ads, word of mouth, and flyers [18]. Lastly, in a cohort description paper among the first 10,000 participants in the Apple Women's Health Study, multiple digital recruitment modalities were used [19].

The objective of this study was to assess the feasibility of recruiting a geographically diverse sample of women to an epidemiological study of OM health via digital recruitment. Feasibility and diversity were assessed via click and participation rates, geographic location, BMI, smoking status, and other demographic information about study participants.

## Methods

### OM Health Study

The OM Health Study is the umbrella study that includes the OM Health Pilot Study and the current OM Global Health Study. The OM Health Study was designed to advance our understanding of factors that promote menstrual health and awareness. Details regarding its health and lifestyle survey

instrument and pilot launch within the United States have been previously described [20,21].

### Ethical Considerations

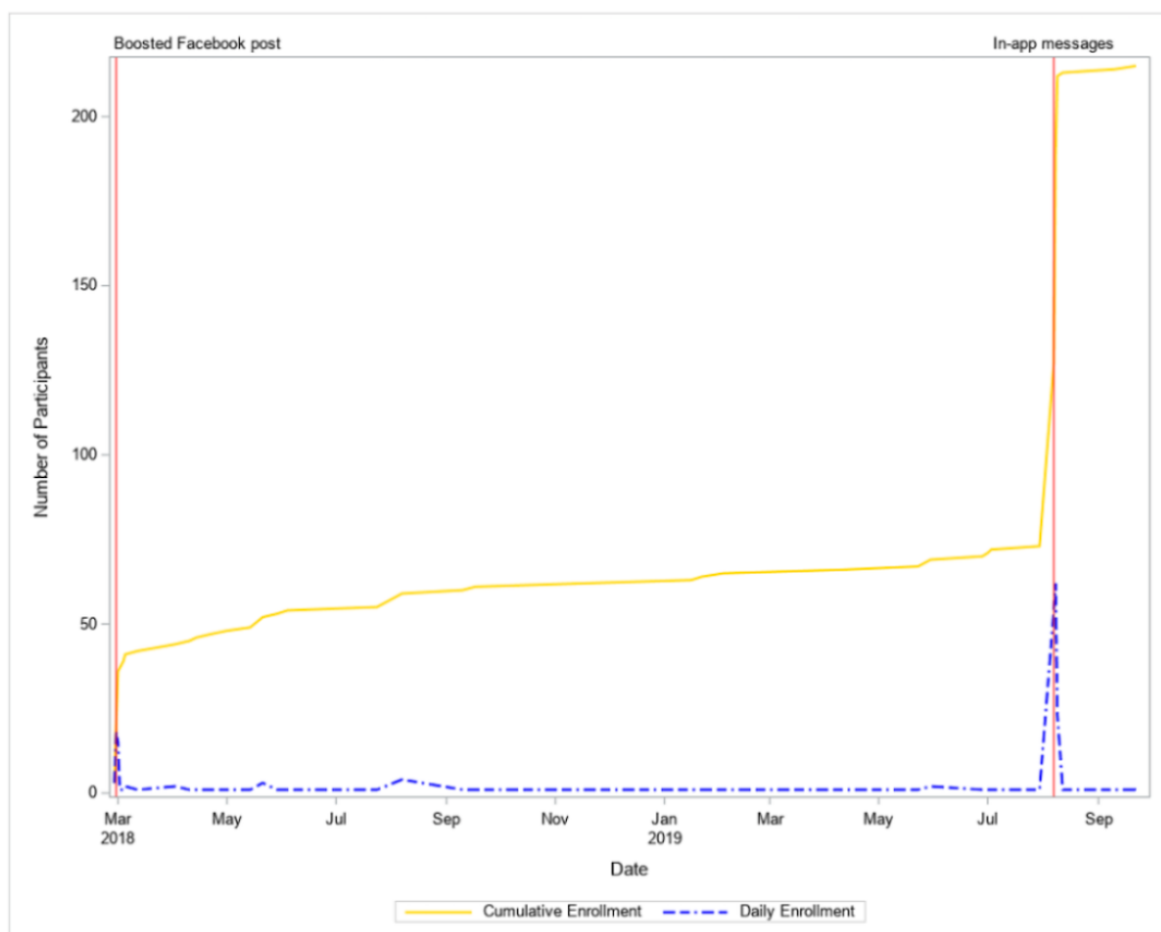
Participants provided informed consent to take part in the study. The web-based consent form included questions used to determine a participant's interest in the current study, follow-up surveys, future studies, and contributing biospecimens. The parent study was approved by the Boston University Medical Campus Institutional Review Board (ID: IRB H-35075). De-identified study data were approved for analysis by the Harvard University Institutional Review Board (ID: IRB20-0638).

### Participant Recruitment

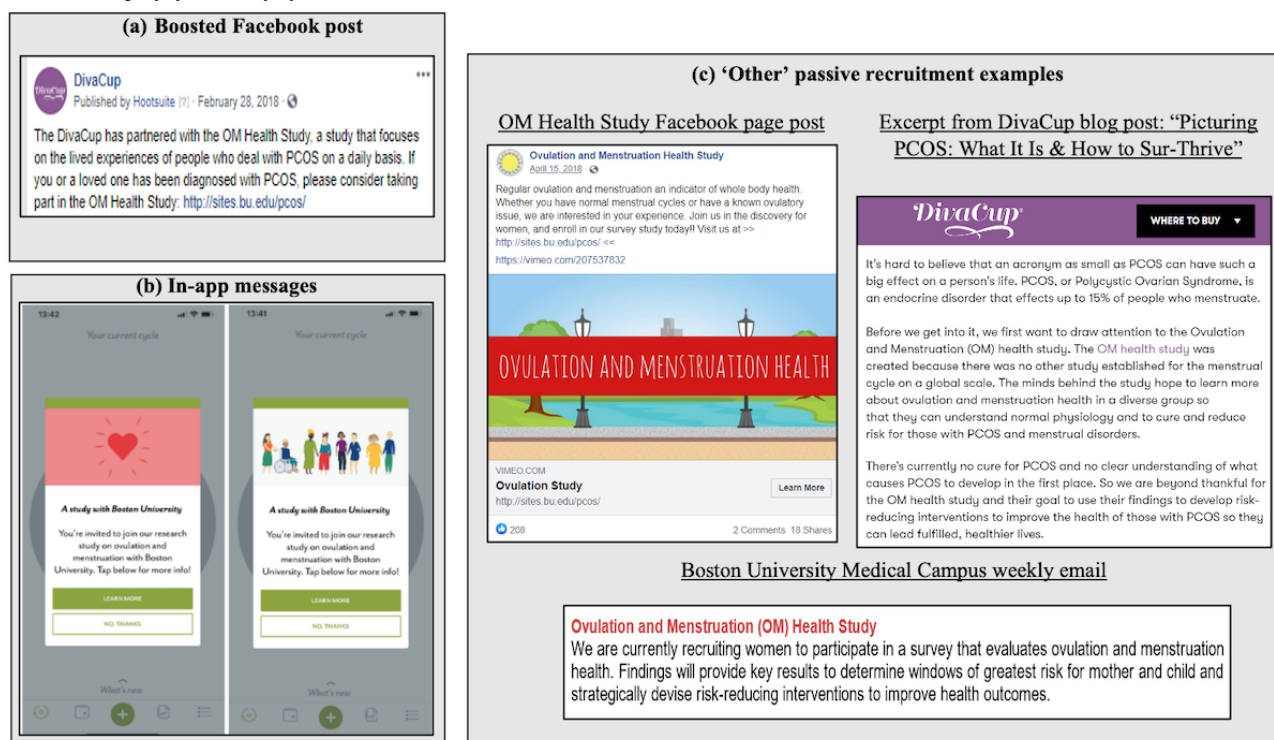
A timeline of participant recruitment into the OM Global Health Study is displayed in Figure 1, and materials used to recruit participants are displayed in Figure 2. Participants were recruited online via 3 recruitment streams. The first was a DivaCup-boosted Facebook post including language related to a diagnosis of polycystic ovary syndrome (PCOS; "boosted Facebook post recruitment") that was deployed on the DivaCup Facebook page on February 28, 2018 (Figure 2a). A boosted Facebook post is an ad that is created from posts on a Facebook page and is intended to create new user engagement and increase current user interactions [22]. The second stream comprised

Clue in-app messages ("app recruitment"). In-app messages were sent to a subset of Clue users who were born after 2001, had the app language set to English, and had not received a survey from Clue within the past 60 days. Additionally, this subset comprised Clue users who opened the app during August 6-19, 2019 (Figure 2b). A third stream included other online passive recruitment study materials between February 27, 2018, and January 24, 2020 (Figure 2c). The other stream ("other passive recruitment") consisted of posts on the OM Global Health Study's social media accounts (LinkedIn, Twitter, and Facebook); DivaCup's social media accounts, blog, or newsletter; or the Boston University Medical Campus weekly email blast that is sent to faculty, staff, and students. Passively recruited individuals were those who engaged with the other passive recruitment materials and those who may have encountered the study website on their own. To be eligible for the OM Global Health Study, individuals were required to be aged 18-44 years, not be currently pregnant, have no history of chemotherapy radiation or surgical menopause, and have a valid email address. The recruitment approach was split into 2 separate communications. Participants were given the link to the Research Electronic Data Capture (REDCap; Vanderbilt University)-administered survey and went through the consent and screener first. If they were eligible, they then received an email with an individualized link to the survey to continue their participation in the study.

**Figure 1.** Timeline of digital recruitment. Note: Vertical red lines denote when active recruitment events (boosted Facebook post and in-app messages) occurred.



**Figure 2.** Digital recruitment materials: (a) boosted Facebook post, (b) in-app messages messaging, and (c) examples from other passive recruitment methods. PCOS: polycystic ovary syndrome.



## Recruitment and Participation Metrics

Click rate metrics were calculated to quantify the interest of potential participants for the OM Global Health Study website from the discrete recruitment events (ie, DivaCup and Clue). Click rate metrics were defined for the boosted DivaCup Facebook post and Clue in-app messages as the number of clicks per the number of ad views as reported by Facebook metrics and the number of clicks per in-app message views as reported by Braze customer engagement software for Clue, respectively.

Overall participation rates were calculated and defined as the number of completed surveys per number of completed consents and eligibility screeners. Survey completion was defined as finishing the pregnancy and birth history section of the OM Global Health Study questionnaire. The completion rate was defined as survey completion per number of initiated surveys. Initiated consents attributed to nonhuman engagement were excluded. We defined nonhuman engagement as survey forms that appeared to be initiated by a web app security scanner. Geographic, demographic, and health-related characteristics of participants who started the survey were evaluated using SAS 9.4 (SAS Institute).

## Results

### Recruitment and Enrollment

The click rate was 18.9% (19,700 clicks/104,000 ad views) via the Clue app and 1.6% via the boosted Facebook post (340

clicks/21,400 ad views). Overall, 215 individuals started the survey, of whom 140 (65.1%), 39 (18.1%), and 36 (16.8%) were recruited via the app, the boosted Facebook post, and other passive recruitment methods, respectively (Table 1). The first participant enrolled on February 27, 2018, and the last participant enrolled on September 22, 2019. Upticks in cumulative enrollment coincided with the deployment of the boosted Facebook post and in-app messages (Figure 1). The survey completion rate was 92.1% (198 completed per 215 started).

A total of 1466 consents and 856 eligibility screeners were completed, resulting in 444 consented and eligible individuals. Among those screened for eligibility (N=856), individuals were excluded if they were under 18 years of age (n=355, 41.5%), did not provide an email address (n=19, 2.2%), or were no longer menstruating (n=38, 4.4%). All those excluded because they were underage were recruited via the app. Of those who were recruited via the app (N=140) and provided a reason for nonparticipation in the eligibility screener (n=69, 49.3%), 64 (92.8%) reported the reason was their age and 52 (75.4%) specifically stated they were under the age of 18 years. The highest number of consents (n=2750, 46.2%) was initiated by an internet bot started on September 7, 2019; none of these progressed past the eligibility screener and were thus unable to start and complete the survey. Thus, the participation rate for the survey was 44.6% (198/444).

**Table 1.** Derivation of final study population and recruitment metrics by recruitment stream.

Metrics	Overall, N	Boosted Facebook post	In-app messages	Other passive recruitment
Date(s) of recruitment	February 27, 2018-January 24, 2020	February 28, 2018	August 6, 2019- August 19, 2019	February 27, 2018-January 24, 2020
Ad views, n (%)	125,400 (100)	21,400 (17.2)	104,000 (82.8)	N/A <sup>a</sup>
Ad clicks, n (%)	20,110 (100)	340 (1.7)	19,770 (98.3)	N/A
<b>Consent, n (%)</b>				
Excluded: internet bot	2750 (100)	N/A	N/A	N/A
Initiated	1755 (100)	74 (4.2)	1598 (91.1)	83 (4.7)
Completed	1446 (100)	73 (5.1)	1293 (89.4)	80 (5.5)
<b>Eligibility screener, n (%)</b>				
Excluded: under 18 years	355 (100)	0	355 (100)	0
Excluded: no longer menstruating	38 (100)	20 (52.6)	2 (5.3)	16 (42.1)
Excluded: email address not provided	19 (100)	0	19 (100)	0
Completed	856 (100)	68 (7.9)	724 (84.6)	64 (7.5)
Eligible and consented	444 (100)	48 (10.8)	348 (78.4)	48 (10.8)
<b>OM<sup>b</sup> Global Health Survey, n (%)</b>				
Started	215 (100)	39 (18.1)	140 (65.1)	36 (16.8)
Completed	198 (100)	37 (18.7)	128 (64.6)	33 (16.7)
<b>Recruitment metrics</b>				
Click rate (clicks/views)	N/A	1.6% (340/21,400)	18.9% (19,700/104,000)	N/A
Survey completion rate (#completed surveys/#eligible and consented)	44.6% (198/444)	77.1% (37/48)	36.8% (128/348)	68.8% (33/48)

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

<sup>b</sup>OM: ovulation and menstruation.

## Participant Characteristics

The average age of the 215 participants who started the survey was 21.8 (SD 6.1) years. The average age of the participants was 28.8 (SD 6.3) years for the boosted Facebook post and 28.0 (SD 5.8) years for other passive recruitment. Due to the programming misspecification in the Clue app, only participants who reported that they were aged 18 years were notified about the study; thus, the average age was 18.3 (SD 1.1) years among Clue recruits (Table 2). Approximately, 119 (60%) of the 198 participants who started and completed the survey were 18 years old.

In terms of racial/ethnic diversity, 147 (70.7%) of the participants overall identified as non-Hispanic White, 35 (16.8%) more than 1 race/ethnicity, 10 (4.8%) Hispanic, 6 (2.9%) Asian, 4 (1.9%) Black, 4 (1.9%) Middle Eastern, and 2 (1.0%) other race/ethnicity. Overall, 123 (59.4%) of the participants had a high school education or less. In addition, 62 (29.8%) participants did not know their household income, 34 (16.4%) preferred not to answer, and the remaining majority

(n=51, 24.5%) fell within the US \$25,000-\$74,999 income range, while 33 (15.9%) reported a household income of US \$100,000 or more. Regarding health-related characteristics, the prevalence of smoking at least 100 cigarettes over their lifetime was 11.3% (n=22), and 31 (15.2%) and 49 (24.0%) of the participants were classified as overweight (BMI=25-29.9 kg/m<sup>2</sup>) and obese (BMI≥30.0 kg/m<sup>2</sup>), respectively. Overall, 104 (51.0%) reported having ever used a hormonal contraceptive, and 87 (43.3%) rated their health as good. There were no notable differences in demographic characteristics by survey completion status (Multimedia Appendix 1, Table S1). The prevalence of health characteristics is presented in Multimedia Appendix 1 (Table S2), and that of health characteristics among those 18 years old is presented in Multimedia Appendix 1 (Table S3). Of note, prevalence within the entire cohort versus those 18 years old was 22.7% and 3.4% for PCOS, 18.7% and 12.7% for gastroesophageal reflux disease (GERD), 21.2% and 23.7% for eating disorders, 3.5% and 3.4% for diabetes, 38.4% and 33.9% for depression, and 44.9% and 46.6% for anxiety, respectively.

**Table 2.** Demographic and health-related characteristics of participants, overall and by recruitment stream (N=215).

Characteristics	Overall (N=215) <sup>a</sup>	Boosted Facebook post (n=39) <sup>b</sup>	In-app messages (n=140) <sup>c</sup>	Other passive recruitment (n=36) <sup>d</sup>
Age (years), mean (SD; range)	21.8 (SD 6.1; 18-44)	28.8 (SD 6.3; 18-44)	18.3 (SD 1.1; 18-26)	28.0 (SD 5.8; 18-43)
<b>Residence<sup>e</sup>, n (%)</b>				
Asia	8 (3.9)	0	6 (4.5)	2 (5.7)
Australia	4 (1.9)	0	4 (3.0)	0
Europe	46 (22.3)	2 (5.4)	44 (32.8)	0
North America (outside the United States)	29 (14.1)	12 (32.4)	15 (11.2)	2 (5.7)
South America	4 (1.9)	1 (2.7)	3 (2.2)	0
United States	115 (55.8)	22 (59.5)	62 (46.3)	31 (88.6)
<b>Race/ethnicity, n (%)</b>				
White (non-Hispanic)	147 (70.7)	32 (84.2)	91 (67.4)	24 (68.6)
Latina/Hispanic	10 (4.8)	0	8 (5.9)	2 (5.7)
Black/African American (non-Hispanic)	4 (1.9)	0	2 (1.5)	2 (5.7)
Asian	6 (2.9)	0	4 (3.0)	2 (5.7)
Middle Eastern	4 (1.9)	0	4 (3.0)	0
Other race/ethnicity	2 (1.0)	0	2 (1.5)	0
More than 1 race/ethnicity	35 (16.8)	6 (15.8)	24 (17.8)	5 (14.3)
<b>Educational attainment, n (%)</b>				
High school graduate/General Educational Development (GED) or less	123 (59.4)	2 (5.3)	120 (89.6)	0 (0.0)
Some college or 2-year degree	26 (12.6)	18 (47.4)	6 (4.5)	1 (2.9)
4-year college graduate	33 (15.9)	13 (34.2)	6 (4.5)	14 (40.0)
More than 4-year college degree	25 (12.1)	5 (13.2)	2 (1.5)	18 (51.4)
<b>Total annual household income (US \$), n (%)</b>				
Below 25,000	16 (7.7)	4 (10.5)	9 (6.7)	3 (8.6)
25,000-49,999	24 (11.5)	8 (21.1)	9 (6.7)	7 (20.0)
50,000-74,999	27 (13.0)	6 (15.8)	11 (8.2)	10 (28.6)
75,000-99,999	12 (5.8)	2 (5.3)	7 (5.2)	3 (8.6)
100,000 or more	33 (15.9)	10 (26.3)	17 (12.6)	6 (17.1)
Prefer not to answer	34 (16.4)	5 (13.2)	28 (20.7)	1 (2.9)
Do not know	62 (29.8)	3 (7.9)	54 (40.0)	5 (14.3)
Smoked at least 100 cigarettes over lifetime, n (%)	22 (11.3)	8 (21.6)	10 (8.0)	4 (12.1)
<b>BMI (kg/m<sup>2</sup>), n (%)</b>				
Underweight (<18.5)	17 (8.3)	0	17 (13.0)	0
Normal weight (18.5-24.9)	107 (52.5)	4 (10.8)	84 (64.1)	19 (52.8)
Overweight (25.0-29.9)	31 (15.2)	5 (13.5)	19 (14.5)	10 (27.8)
Obese (≥30.0)	49 (24.0)	28 (75.7)	11 (8.4)	7 (19.4)
Hormonal contraceptives use ever, n (%)	104 (51.0)	35 (94.6)	38 (28.8)	31 (88.6)
<b>Participant's rating of current health, n (%)</b>				
Excellent	22 (11.0)	1 (2.7)	14 (10.8)	7 (20.6)
Very good	66 (32.8)	11 (29.7)	40 (30.8)	15 (44.1)

Characteristics	Overall (N=215) <sup>a</sup>	Boosted Facebook post (n=39) <sup>b</sup>	In-app messages (n=140) <sup>c</sup>	Other passive recruitment (n=36) <sup>d</sup>
Good	87 (43.3)	19 (51.4)	58 (44.6)	10 (29.4)
Fair	24 (11.9)	6 (16.2)	16 (12.3)	2 (5.9)
Poor	2 (1.0)	0 (0.0)	2 (1.5)	0 (0.0)

<sup>a</sup>n (% complete): age, 215 (100%); residence, 206 (95.8%); race/ethnicity, 208 (96.7%); education, 207 (96.3%); income, 208 (96.7%); smoking status, 195 (90.7%); BMI, 204 (94.9%); hormonal contraceptive, 204 (94.9%); self-rated current health, 201 (93.5%); “complete” defined as pregnancy birth history complete in the questionnaire.

<sup>b</sup>n (% complete): age, 39 (100%); residence, 37 (94.9%); race/ethnicity, 38 (97.4%); education, 38 (97.4%); income, 38 (97.4%); smoking status, 37 (94.9%); BMI, 37 (94.9%); hormonal contraceptive, 37 (94.9%); self-rated current health, 37 (94.9%).

<sup>c</sup>n (% complete): age, 140 (100%); residence, 134 (95.7%); race/ethnicity, 135 (96.4%); education, 134 (95.7%); income, 135 (96.4%); smoking status, 125 (89.3%); BMI, 131 (93.6%); hormonal contraceptive, 132 (94.3%); self-rated current health, 130 (92.9%).

<sup>d</sup>n (% complete): age, 36 (100%); residence, 35 (97.2%); race/ethnicity, 35 (97.2%); education, 35 (97.2%); income, 35 (97.2%); smoking status, 33 (91.7%); BMI, 36 (100%); hormonal contraceptive, 35 (97.2%); self-rated current health, 34 (94.4%).

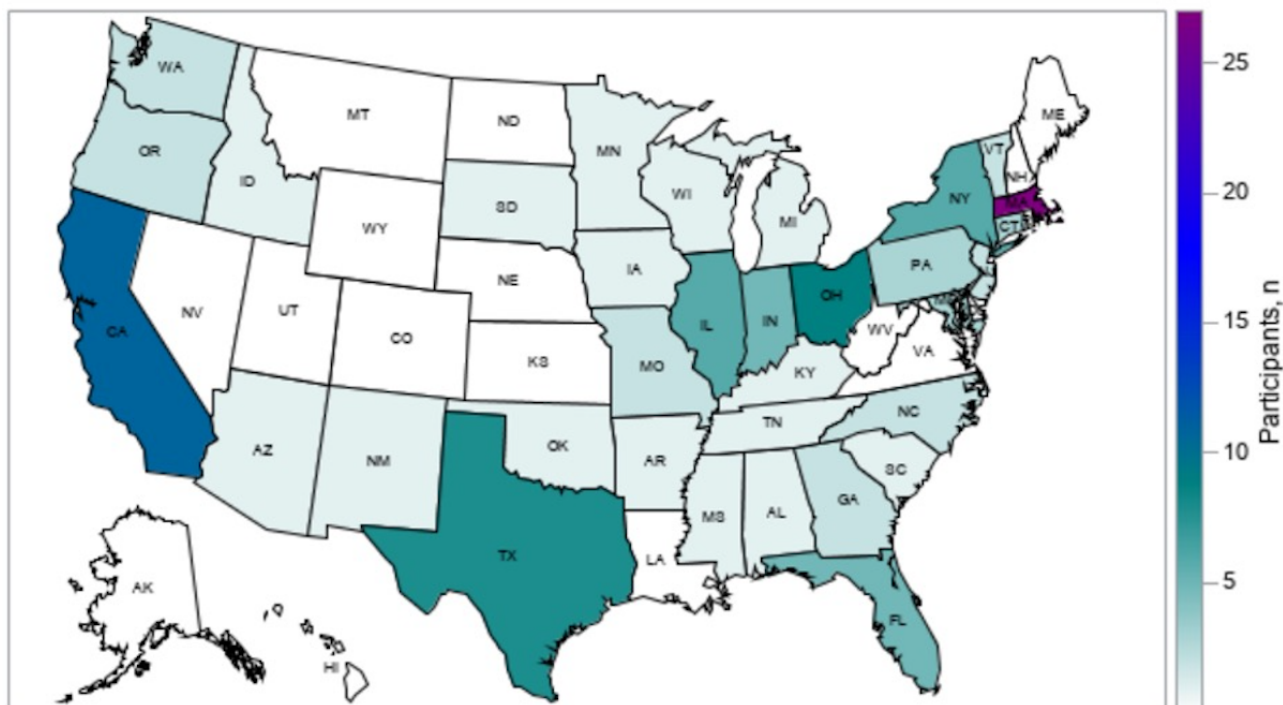
<sup>e</sup>“Residence” defined as country of birth reported and reporting not living in the United States.

### Geographic Diversity

Overall, 115 (55.8%) of the 206 participants recruited via different modalities who completed the residence question resided in the United States, 46 (22.3%) in Europe, 29 (14.1%) outside the United States in North America, and approximately 16 (7.8%) in other parts of the world (Table 2). Recruitment via the Clue app (Germany) produced the most geographically diverse stream: 77 (57.5%) of the 134 participants recruited via the app who completed the residence question resided in North America, 44 (32.8%) in Europe, 6 (4.5%) in Asia, 4 (3.0%) in

Australia, and 3 (2.2%) in South America. The boosted Facebook post through DivaCup (Canada) had the greatest proportion of participants who lived in parts of North America outside the United States (12/37, 32.4%). Most participants recruited via other passive recruitment methods (31/35, 88.6%) resided in the United States. Participants who were recruited within the United States, through all the modalities, resided in 33 different states, with over half located in 6 states: Massachusetts (n=27, 23.5%), California (n=11, 9.6%), Ohio (n=9, 7.8%), Texas (n=8, 7.0%), Illinois (n=6, 5.2%), and New York (n=6, 5.2%); see Figure 3.

**Figure 3.** Geographic distribution of participants within the United States.



### Technical Recruitment Errors

Two major errors occurred during the study period. One appeared to be due to an internet bot, and the second was a programming misspecification in the Clue app. Although these can be considered procedural breakdowns, screening of potential

enrollees ensured eligibility requirements were met for those who did participate in the study. The internet bot appeared to result from a hacking attempt into the REDcap survey. This resulted in 2750 consent forms started on 1 day, with no further progression onto the study forms. These inputs were excluded

from the study. The second procedural breakdown involved the age of those recruited via the Clue app. This was attributable to a programming misspecification for targeting the in-app message, which set the target birth year as greater than 2001 (instead of earlier than 2001). All these individuals were excluded from the study since they were under 18 years of age.

## Discussion

### Principal Findings

Digital recruitment via in-app messaging, a boosted Facebook post, and other passive recruitment methods was determined to be feasible. Previous studies have demonstrated that internet-based recruitment is cost-effective and typically increases cohort diversity, although it is limited geographically to specific regions, as seen in Fenner et al [17], Wise et al [18], and Mahaligaiah et al [19]. Moreover, previous studies have illustrated that diversity has increased in digital recruitment studies. Through the use of smartphone-based technology specifically, as was done in the MyHeart Counts study [23] and the Apple Women's Health Study [19], our descriptive paper elaborated on the utility of these methods and, as detailed next, highlighted various limitations of each modality.

### Click and Participation Rates

We had a 44.6% overall participation rate, which is slightly above a previous health study that recruited women aged 16-25 years via a Facebook ad (24.5%) [17]. Although in-app recruitment had a higher click rate compared to the boosted Facebook post (19% vs 1.6%), the interpretation of these results was limited due to the programming misspecification that occurred upon initial release of the messaging. In addition, 75% of women from the in-app recruitment stream were not able to participate because they were under 18 years of age. Thus, this sample had overrepresentation of 18-year-olds. Future studies to assess age-related health risks will require additional stratification by age.

### Health Characteristics

We found that individuals who chose to participate across all 3 recruitment platforms had similar health characteristics, such as the BMI, smoking, and hormonal contraceptive use. These health-related findings likely reflect the large proportion of participants who were 18 years old, due to the programming misspecification. Thus, these findings may not be generalizable to the intended study population, which would have included larger proportions of individuals across the targeted age range of 18-45 years. Future studies should ensure that such problems do not arise in order to determine how digital platforms can be used to recruit individuals across a wider age range and with more variable health characteristics. Similarly, recruitment messaging targeting PCOS in the DivaCup post may oversample women with this disease [24]. Women using mobile menstrual tracking apps or those motivated to consume reproductive health information online may have differing goals, such as disease management, fertility enhancement, contraception, or health optimization, that may introduce selection bias that could impact measures of association derived from an epidemiological study using these forms of recruitment [5].

### Geographic Diversity

Diva International Inc. targets its advertisements to women primarily in Canada, the United States, Mexico, and Australia [25], while BioWink GmbH, which is based in Germany, is used in 190 countries worldwide [15]. The geographic reach of these companies is reflected in the geographic spread of our participants (ie, 44% were international). Additionally, participants recruited in the OM Global Health Study were more geographically diverse within the United States (33 states) compared to the OM Pilot Study (20 states) that relied primarily on in-person recruitment strategies [21]. Future epidemiological studies utilizing digital recruitment may benefit from partnering with similar companies with a wide geographic user base to target diverse participants for their research.

Additionally, in-app recruitment via Clue was able to not only reach a large audience (104,000 in-app message views) but also was rapid (in-app messaging occurred over a 2-week period) and inexpensive compared to traditional health research recruitment. Thus, in-app recruitment may be a cost-effective strategy that may enable health research in an increasingly strained funding environment.

### Limitations

Our study leveraged the consumer base of 2 major companies, Diva International Inc. and BioWink GmbH, and demonstrated the feasibility of using the platforms of these companies to recruit participants. Our results also demonstrated an opportunity to use mobile apps for recruitment of younger participants for research. Despite the widespread access to the internet and ownership of smartphones, some potential limitations of our study are the lack of ownership of electronic devices, such as smartphones and computers; digital illiteracy; and language barriers. Another major limitation of our study is the average age of those recruited via Clue due primarily to a programming misspecification. Although it would have been useful to see the age distribution of participants with targeting including those older than 18 years, our results show that younger women are interested in participating in research. Of those individuals who provided a reason for declining participation, 93% (n=64) stated that they were aged less than 18 years and 94% of ineligible participants (n=355) were deemed so due to age. Similarly, a prior study using the menstrual tracking app Flo to assess cycle variability found that approximately half of their participants were between the ages of 18 and 24 years [26]. These streams may present an opportunity for recruiting teenage women into health research. More importantly, this finding demonstrates that adequate study staff will be needed to reach out to participants and confirm eligibility, as was done by Fenner et al [17]. In addition to the protection of minors, it will be important to consider how demographic and health indicators are influenced by age. For example, younger generations tend to self-rate their health better than older generations [27].

Lastly, this study provides lessons for optimizing the use of digital tools in future research, including avoiding human and nonhuman sources of procedural breakdowns and improving information technology management. With traditional in-person recruitment, errors can be more easily identified than with digital recruitment. For example, it is simple for research assistants



working in a clinic to identify recruitment errors, such as approaching participants outside a specified age and correcting them in real time. Additionally, hacking attempts by internet bots are specific to a digital space. Public surveys, especially those distributed via social media, are susceptible to hacking attempts; thus, implementing the “endCAPTCHA” module in REDcap may prevent this issue as it would provide insurance against bots inputting data as it serves as a human verification tool. To increase the utility of recruitment in the digital space, we recommend careful assessment of the programming process in real time in order to quickly identify and correct any anomalies in the resulting data. Server capacities also introduced limitations to our study. Initially, the consent form took up to 1-2 minutes to load, which may have resulted in participant drop-off before women were able to complete the eligibility and screening forms. Not only did the consent form load slowly, but when opening the link from an email, loading time varied, which may have also contributed to the low number of women enrolling in the survey. This may also have resulted in the exclusion of individuals with low income who tend to rely on a free or inexpensive Wi-Fi connection, exacerbating the underrepresentation of low-income women in health research

[28]. Another issue would be language barriers in accessing our survey. As 1 of the inclusion criteria was being able to read English, we may have limited the scope of individuals who may have otherwise participated in this study.

## Conclusion

The broad scope of digital recruitment in this study allowed for more geographically, racially/ethnically, and age diverse participants than our OM Pilot Study, which relied on primarily in-person recruitment strategies. This study demonstrated the utility of a digital recruitment approach to successfully recruit participants worldwide. In-app recruitment resulted in the greatest number of surveys completed and contributed the most geographic diversity. Recruitment via the boosted Facebook post and other passive recruitment methods helped us further target specific audiences, such as those interested in participating due to general interest in women’s health. Procedural breakdowns also demonstrated the possible challenges in engaging with various digital platforms during research studies, while highlighting an opportunity to engage premenarchal- and menarchal-aged women in health research via app recruitment. Our findings and limitations may be useful to inform future epidemiological studies implementing digital recruitment.

## Acknowledgments

The authors would like to acknowledge the support of the DivaCup team for posting about the study, J Jojo Cheng for his foundational work on the ovulation and menstruation (OM) Health Study, and Kimberly Dukes for answering questions about the Biostatistics and Epidemiology Data Analytics Center (BEDAC) at the Boston University Medical Campus. AW was supported by the New York Medical College (NYMC) School of Medicine Medical Student Affairs Summer Fellowship, and KJL was funded in part by the Boston University School of Public Health Career Catalyst pilot award funded by the Idea Hub and the Robert F Meenan Faculty Support Fund. The OM Health Study was funded by a Reproductive Scientist Development Program Seed Grant 2016, and KP was supported by an environmental epidemiology grant (#T32 ES007069). All authors contributed to the conception of this project. ER, AdFV, AS, and SM contributed to establishing the Clue app recruitment. KP, VF, and MW conducted the data analysis. KP and ER wrote the initial draft, and all authors contributed to interpreting, revising, and reviewing the manuscript. KJ contributed to the tables.

## Conflicts of Interest

AS was employed by Clue by BioWink GmbH at the time of the study. All other authors report no conflict of interest.

## Multimedia Appendix 1

Demographic characteristics by survey completion status and the prevalence of health characteristics.

[DOCX File, 30 KB - [formative\\_v6i9e39046\\_app1.docx](#) ]

## References

1. Pew Research Center. Internet/Broadband Fact Sheet. URL: <https://www.pewresearch.org/internet/fact-sheet/internet-broadband/?menuItem=d5edf003-5858-4269-89c5-f2889ecf7951> [accessed 2022-08-09]
2. Pew Research Center. Mobile Fact Sheet. URL: <https://www.pewresearch.org/internet/fact-sheet/mobile/> [accessed 2022-08-09]
3. Critchley HO, Babayev E, Bulun SE, Clark S, Garcia-Grau I, Gregersen PK, et al. Menstruation: science and society. *Am J Obstet Gynecol* 2020 Nov;223(5):624-664 [FREE Full text] [doi: [10.1016/j.ajog.2020.06.004](https://doi.org/10.1016/j.ajog.2020.06.004)] [Medline: [32707266](https://pubmed.ncbi.nlm.nih.gov/32707266/)]
4. Bradley S, Bacharach E. The 11 Best Period Tracker Apps, According to Ob-Gyns. 2019. URL: <https://www.womenshealthmag.com/health/g26787041/best-period-tracking-apps/> [accessed 2022-08-19]
5. Schantz JS, Fernandez CSP, Anne Marie ZJ. Menstrual cycle tracking applications and the potential for epidemiological research: a comprehensive review of the literature. *Curr Epidemiol Rep* 2021 Mar 20;8(1):9-19 [FREE Full text] [doi: [10.1007/s40471-020-00260-3](https://doi.org/10.1007/s40471-020-00260-3)] [Medline: [34055569](https://pubmed.ncbi.nlm.nih.gov/34055569/)]

6. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. *NPJ Digit Med* 2019 Jun 3;2(1):45 [FREE Full text] [doi: [10.1038/s41746-019-0121-1](https://doi.org/10.1038/s41746-019-0121-1)] [Medline: [31304391](https://pubmed.ncbi.nlm.nih.gov/31304391/)]
7. Lane TS, Armin J, Gordon JS. Online recruitment methods for web-based and mobile health studies: a review of the literature. *J Med Internet Res* 2015 Jul 22;17(7):e183 [FREE Full text] [doi: [10.2196/jmir.4359](https://doi.org/10.2196/jmir.4359)] [Medline: [26202991](https://pubmed.ncbi.nlm.nih.gov/26202991/)]
8. Cowie JM, Gurney ME. The use of Facebook advertising to recruit healthy elderly people for a clinical trial: baseline metrics. *JMIR Res Protoc* 2018 Jan 24;7(1):e20 [FREE Full text] [doi: [10.2196/resprot.7918](https://doi.org/10.2196/resprot.7918)] [Medline: [29367186](https://pubmed.ncbi.nlm.nih.gov/29367186/)]
9. Akers L, Gordon JS. Using Facebook for large-scale online randomized clinical trial recruitment: effective advertising strategies. *J Med Internet Res* 2018 Nov 08;20(11):e290 [FREE Full text] [doi: [10.2196/jmir.9372](https://doi.org/10.2196/jmir.9372)] [Medline: [30409765](https://pubmed.ncbi.nlm.nih.gov/30409765/)]
10. Wozney L, Turner K, Rose-Davis B, McGrath PJ. Facebook ads to the rescue? Recruiting a hard to reach population into an Internet-based behavioral health intervention trial. *Internet Interv* 2019 Sep;17:100246 [FREE Full text] [doi: [10.1016/j.invent.2019.100246](https://doi.org/10.1016/j.invent.2019.100246)] [Medline: [31080751](https://pubmed.ncbi.nlm.nih.gov/31080751/)]
11. ACOG Committee opinion no. 649: racial and ethnic disparities in obstetrics and gynecology. *Obstet Gynecol* 2015 Dec;126(6):e130-e134. [doi: [10.1097/aog.0000000000001213](https://doi.org/10.1097/aog.0000000000001213)]
12. Baquet CR, Commiskey P, Daniel Mullins C, Mishra SI. Recruitment and participation in clinical trials: socio-demographic, rural/urban, and health care access predictors. *Cancer Detect Prev* 2006 Jan;30(1):24-33 [FREE Full text] [doi: [10.1016/j.cdp.2005.12.001](https://doi.org/10.1016/j.cdp.2005.12.001)] [Medline: [16495020](https://pubmed.ncbi.nlm.nih.gov/16495020/)]
13. Statista Research Department. Feminine Hygiene Market - Statistics & Facts. URL: <https://www.statista.com/topics/4889/feminine-hygiene-market/#dossierSummary> [accessed 2022-08-18]
14. Grand View Research. Women's Health App Market Size, Share & Trends Analysis Report By Type (Fitness & Nutrition, Pregnancy Tracking & Postpartum Care, Menopause), By Region (North America, Europe, APAC, Latin America, MEA), And Segment Forecasts, 2021 - 2028. URL: <https://www.grandviewresearch.com/industry-analysis/womens-health-app-market> [accessed 2022-08-18]
15. Clue by Biowink GmbH. The App to Track Your Period, and So Much More. URL: <https://helloclude.com/period-tracker-app> [accessed 2022-08-19]
16. Saunders A. DivaCup International Inc Media Kit. 2020. URL: [https://divainternational.ca/wp-content/uploads/2020/07/July-20-English-Media-Kit-2020\\_web.pdf](https://divainternational.ca/wp-content/uploads/2020/07/July-20-English-Media-Kit-2020_web.pdf) [accessed 2022-08-19]
17. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. *J Med Internet Res* 2012 Feb 01;14(1):e20 [FREE Full text] [doi: [10.2196/jmir.1978](https://doi.org/10.2196/jmir.1978)] [Medline: [22297093](https://pubmed.ncbi.nlm.nih.gov/22297093/)]
18. Wise LA, Rothman KJ, Mikkelsen EM, Stanford JB, Wesselink AK, McKinnon C, et al. Design and conduct of an internet-based preconception cohort study in North America: Pregnancy Study Online. *Paediatr Perinat Epidemiol* 2015 Jul;29(4):360-371 [FREE Full text] [doi: [10.1111/ppe.12201](https://doi.org/10.1111/ppe.12201)] [Medline: [26114445](https://pubmed.ncbi.nlm.nih.gov/26114445/)]
19. Mahalingaiah S, Fruh V, Rodriguez E, Konanki SC, Onnela J, de Figueiredo Veiga A, et al. Design and methods of the Apple Women's Health Study: a digital longitudinal cohort study. *Am J Obstet Gynecol* 2022 Apr;226(4):545.e1-545.e29. [doi: [10.1016/j.ajog.2021.09.041](https://doi.org/10.1016/j.ajog.2021.09.041)] [Medline: [34610322](https://pubmed.ncbi.nlm.nih.gov/34610322/)]
20. Mahalingaiah S, Cosenza C, Cheng JJ, Rodriguez E, Aschengrau A. Cognitive testing of a survey instrument for self-assessed menstrual cycle characteristics and androgen excess. *Fertil Res Pract* 2020 Nov 09;6(1):19 [FREE Full text] [doi: [10.1186/s40738-020-00088-x](https://doi.org/10.1186/s40738-020-00088-x)] [Medline: [33292647](https://pubmed.ncbi.nlm.nih.gov/33292647/)]
21. Mahalingaiah S, Cheng JJ, Winter MR, Rodriguez E, Fruh V, Williams A, et al. Multimodal recruitment to study ovulation and menstruation health: internet-based survey pilot study. *J Med Internet Res* 2021 Apr 16;23(4):e24716 [FREE Full text] [doi: [10.2196/24716](https://doi.org/10.2196/24716)] [Medline: [33861203](https://pubmed.ncbi.nlm.nih.gov/33861203/)]
22. Briggs-Gill J, Best E. What Happened? Facebook Will Be Removing the 28-Day Attribution Option and Instead, Defaulting to 7-Day Attribution. 2021. URL: <https://www.seerinteractive.com/blog/facebook-attribution-changing-2021/#:~:text=Facebook%20will%20be%20removing%20the.ad%20and%20prior%20to%20converting> [accessed 2022-08-19]
23. 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](https://doi.org/10.1001/jamacardio.2016.4395)] [Medline: [27973671](https://pubmed.ncbi.nlm.nih.gov/27973671/)]
24. Khazaal Y, van Singer M, Chatton A, Achab S, Zullino D, Rothen S, et al. Does self-selection affect samples' representativeness in online surveys? An investigation in online video game research. *J Med Internet Res* 2014 Jul 07;16(7):e164 [FREE Full text] [doi: [10.2196/jmir.2759](https://doi.org/10.2196/jmir.2759)] [Medline: [25001007](https://pubmed.ncbi.nlm.nih.gov/25001007/)]
25. Lombardo C. DivaCup Aims to Start a Period Revolution. URL: <https://strategyonline.ca/2019/02/28/divacup-aims-to-start-a-period-revolution/> [accessed 2022-08-18]
26. Grieger JA, Norman RJ. Menstrual cycle length and patterns in a global cohort of women using a mobile phone app: retrospective cohort study. *J Med Internet Res* 2020 Jun 24;22(6):e17109 [FREE Full text] [doi: [10.2196/17109](https://doi.org/10.2196/17109)] [Medline: [32442161](https://pubmed.ncbi.nlm.nih.gov/32442161/)]
27. Aguilar-Palacio I, Gil-Lacruz AI, Sánchez-Recio R, Rabanaque MJ. Self-rated health in Europe and its determinants: does generation matter? *Int J Public Health* 2018 Mar 30;63(2):223-232. [doi: [10.1007/s00038-018-1079-5](https://doi.org/10.1007/s00038-018-1079-5)] [Medline: [29383383](https://pubmed.ncbi.nlm.nih.gov/29383383/)]

28. Chilukuri N, West M, Henderson JL, Lawson S, Ehsanipoor R, Costigan K, et al. Information and communication technology use among low-income pregnant and postpartum women by race and ethnicity: a cross-sectional study. *J Med Internet Res* 2015 Jul 03;17(7):e163 [[FREE Full text](#)] [doi: [10.2196/jmir.3916](https://doi.org/10.2196/jmir.3916)] [Medline: [26142162](https://pubmed.ncbi.nlm.nih.gov/26142162/)]

## Abbreviations

**OM:** ovulation and menstruation

**PCOS:** polycystic ovary syndrome

*Edited by A Mavragani; submitted 26.04.22; peer-reviewed by G Hassen, H Gandhi; comments to author 03.06.22; revised version received 10.08.22; accepted 15.08.22; published 14.09.22.*

*Please cite as:*

Rodriguez E, Peer K, Fruh V, James K, Williams A, de Figueiredo Veiga A, Winter MR, Shea A, Aschengrau A, Lane KJ, Mahalingaiah S

*Digital Global Recruitment for Women's Health Research: Cross-sectional Study*

*JMIR Form Res* 2022;6(9):e39046

URL: <https://formative.jmir.org/2022/9/e39046>

doi: [10.2196/39046](https://doi.org/10.2196/39046)

PMID: [35969168](https://pubmed.ncbi.nlm.nih.gov/35969168/)

©Erika Rodriguez, Komal Peer, Victoria Fruh, Kaitlyn James, Anna Williams, Alexis de Figueiredo Veiga, Michael R Winter, Amanda Shea, Ann Aschengrau, Kevin J Lane, Shruthi Mahalingaiah. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 14.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Agreement Between Clinically Measured Weight and Self-reported Weight Among Patients With Type 2 Diabetes Through an mHealth Lifestyle Coaching Program in Denmark: Secondary Analysis of a Randomized Controlled Trial

Albi Imeraj<sup>1</sup>; Thomas Bastholm Olesen<sup>2</sup>, MD, PhD; Ditte Hjorth Laursen<sup>3</sup>, MPH, PhD; Jens Søndergaard<sup>1</sup>, MD, PhD; Carl Joakim Brandt<sup>1</sup>, MD, PhD

<sup>1</sup>Research Unit for General Practice, Department of Public Health, University of Southern Denmark, Odense, Denmark

<sup>2</sup>Steno Diabetes Center Odense, Odense University Hospital, Odense, Denmark

<sup>3</sup>Department of Public Health, University of Copenhagen, Copenhagen, Denmark

**Corresponding Author:**

Albi Imeraj

Research Unit for General Practice, Department of Public Health

University of Southern Denmark

J B Winsløws Vej 9A

Odense, 5000

Denmark

Phone: 45 31390187

Email: [aimeraj@health.sdu.dk](mailto:aimeraj@health.sdu.dk)

## Abstract

**Background:** Digital health interventions are increasingly used to handle and promote positive health behaviors. Clinical measures are often used, and a certain precision is essential for digital health interventions to have an effect. Only few studies have compared clinically measured weights with self-reported weights. No study has examined the validity of self-reported weight from a mobile app used in a tailored weight loss intervention.

**Objective:** The aim of this study was to analyze the agreement between clinically measured weight and self-reported weight collected from a mobile health lifestyle coaching program during a 12-month weight loss intervention for obese patients with and without type 2 diabetes. The secondary aim was to investigate the determinants for possible discrepancies between clinically measured and self-reported weights of these patients with different demographic and lifestyle characteristics and achievements of weight loss goals.

**Methods:** Weight registrations were collected from participants (N=104) in a Danish randomized controlled trial examining the effect of a digital lifestyle intervention on weight loss among obese patients with and without type 2 diabetes. Data were collected at baseline and after 6 and 12 months. Self-reported weight was measured at home and registered in the app.

**Results:** Self-reported body weight was lower than the weight measured in the clinic after 6 months by 1.03 kg (95% CI 1.01-1.05;  $P < .001$ ) and after 12 months also by 1.03 kg (95% CI 0.99-1.04;  $P < .001$ ). After 6 months, baseline weight and BMI were associated with a discrepancy of 0.03 kg (95% CI 0.01-0.04;  $P = .01$ ) and 0.09 kg (95% CI 0.02-0.17;  $P = .02$ ) per increment of 1 kg and 1 kg/m<sup>2</sup>, respectively, between clinically measured weight and self-reported weight. Weight change during the first 6 months was also associated with a difference of 0.1 kg (95% CI 0.04-0.01;  $P < .001$ ) per kilogram of difference in weight between clinically measured weight and self-reported weight. Participants who did not achieve the 5% weight loss goal underestimated their weight by 0.79 kg (95% CI 0.34-1.23) at 6 months. After 12 months, only baseline weight was associated with a discrepancy of 0.03 kg (95% CI 0.01-0.05;  $P = .02$ ) per increment of kilogram between clinically measured weight and self-reported weight. None of the other factors showed any significant discrepancy after 12 months.

**Conclusions:** Self-reported weight obtained from mobile health is a valid method for collecting anthropometric measurements.

**Trial Registration:** ClinicalTrials.gov NCT03788915; <https://clinicaltrials.gov/ct2/show/NCT03788915>

(JMIR Form Res 2022;6(9):e40739) doi:[10.2196/40739](https://doi.org/10.2196/40739)

**KEYWORDS**

telemedicine; digital behavioral coaching; lifestyle change; mobile intervention; obesity; diabetes; patient engagement; validation; self-report; body weight

## Introduction

Systematic reviews show that there are several digital health interventions (DHIs) currently that aim to handle and promote positive health behaviors, such as mobile health (mHealth) or web-based interventions [1-6]. DHIs can improve health behavior and weight loss at a reasonable cost [1-6]. Obesity is associated with chronic lifestyle diseases such as type 2 diabetes (T2D) mellitus, cardiovascular disease, and some forms of cancer [1,7]. This is especially the case for T2D, which is strongly correlated to weight gain and obesity. Pathophysiological studies [8,9] indicate that weight loss may normalize glucose control in approximately 50% of patients with T2D. Due to digital advancements, DHIs can now be used to handle and promote positive health behaviors, including self-reporting of weight to track weight loss. Despite this, no clear guidelines or infrastructure have yet been developed for how all these self-reported data should be handled and used in clinical practice. If digital solutions are to be useful, implementation and accessibility of self-reported data are essential. Self-reporting of weight loss is recommended as an effective weight loss strategy and can be performed via different types of DHIs [10]. DHIs are commonly used in both commercial programs and research studies [11,12]. So far, only few studies [13-15] have attempted to evaluate the validity of self-reported weight from web-based and paper-based programs against clinically measured weight, and these studies suggest that self-reported weight may be used as a valid, quick, and cost-effective alternative to clinically measured weight. Furthermore, few studies have reported that the validity of self-reported weight declines with increasing BMI and women tend to underestimate their own weight [13-15]. However, to our knowledge, no study has attempted to investigate the agreement between clinically measured weight and self-reported weight in a mobile app-based lifestyle coaching program. No study has examined whether clinically measured weight and self-reported weight differ (1) with achievement/nonachievement of own weight loss goals and (2) between follow-ups in a 12-month mHealth-based tailored weight loss intervention in a group of overweight people with and without diabetes, where correct weight control is essential.

The primary aim of this study was to determine the agreement between clinically measured weight and self-reported weight collected from an mHealth lifestyle coaching program (long-term Lifestyle change InterVention and mHealth Application [Liva]) during a tailored 12-month weight loss intervention for obese patients with and without T2D. The secondary aim was to investigate the determinants for possible discrepancies between self-reported and clinically measured weights of these patients with different demographic and lifestyle characteristics and achievements of weight loss goals.

## Methods

### Study Design

This study was a secondary analysis and examined the agreement between clinically measured weight and self-reported weight recorded in an mHealth-based solution among intervention participants (N=104) from an open randomized controlled trial (RCT). The control group in the RCT did not have access to the app and therefore had no self-reported weights. We excluded the control group for this study purpose. The RCT examined the effect of a digital lifestyle intervention on weight loss among obese patients with and without T2D. This analysis was conducted in 2 regions in Denmark: the Region of Southern Denmark with 22 municipalities and the Capital Region of Denmark with 28 municipalities. Data were collected from March 2019 to October 2021. All methods are described in further detail in the study protocol [16]. The self-reported weight was collected from the Liva Healthcare mHealth lifestyle coaching program. Patient data included in the study are pseudonymized. Participants granted their consent to make them available for research purposes. Consent was obtained explicitly in the sign-up flow before the use of the app/service.

### Ethics Approval

The RCT was approved by the scientific and ethics committee of the Region of Southern Denmark according to Danish law (approval 18803) and registered on clinicaltrials.gov (NCT03788915).

### Participants and Eligibility Criteria

In each municipality within the participating regions, the participants were recruited through general practitioners and local health centers, the Danish Diabetes Association, and advertisements via social media. The participants registered through the Liva Healthcare app [16]. After registration, a research assistant would contact the participant by phone to make sure that he/she met the following inclusion criteria: (1) BMI of 30-45 kg/m<sup>2</sup>, (2) diagnosed with T2D, and (3) age between 18 and 70 years. The following exclusion criteria were applied: (1) lack of internet access through computer or smartphone, (2) pregnancy or planned pregnancy, and (3) serious or life-threatening disease [16].

### Baseline Meeting and Follow-up Assessment

Participants gave written informed consent and informed the research assistant about their medications at the baseline meeting, and a brief medical examination of the participants was performed subsequently. The medical examination included measurements such as height (measured in centimeters without shoes), weight (without shoes and subtracted 1 kilogram for clothing), and waist and hip circumference (with tape measure around the waist). Weight was measured on a CE-marked high-quality calibrated scale from Tanita Corporation with a capacity of up to 270 kg and weight accuracy of 100 grams.

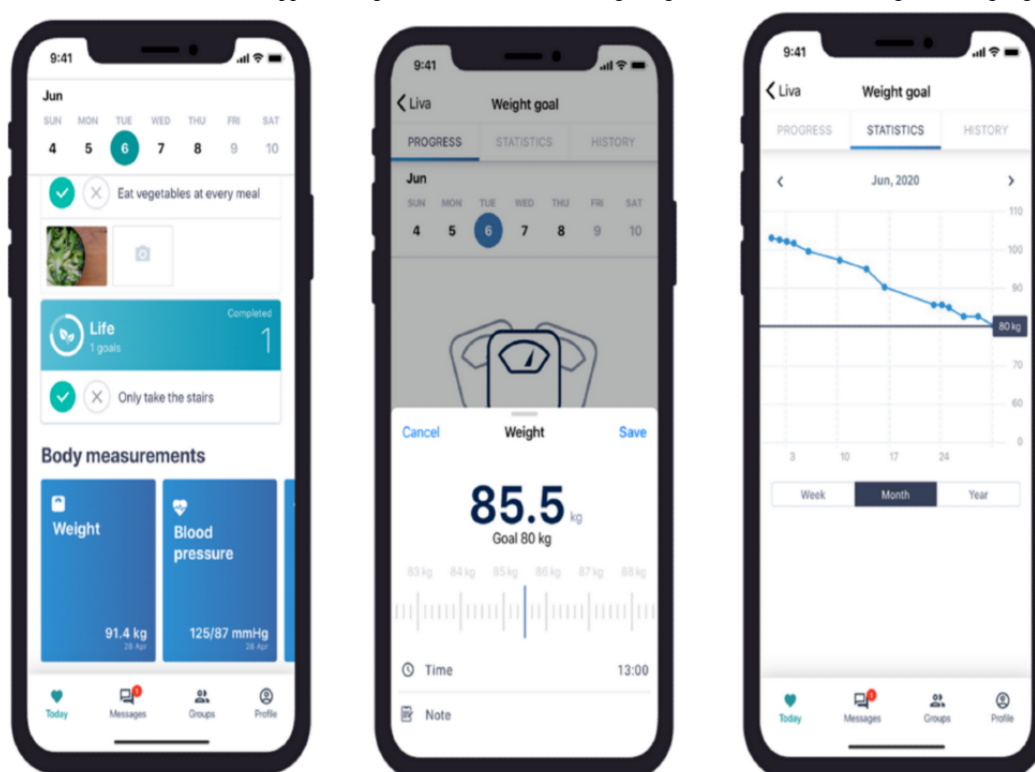
The same measurements were taken at 6 and 12 months of clinical follow-up. As described in the study protocol [16], additional examinations were made but were not included in this study since they were not relevant to our objectives. However, these additional examinations could have an impact on adherence to the intervention.

### Data Collection of Self-reported Weight

The intervention group received access to a lifestyle app/mHealth tool, where they received individual lifestyle coaching, completed daily tasks, and could send remarks or questions directly to the health care professionals (HCPs). The participants could set individual goals using the SMART (specific, measurable, attainable, relevant, timely) model [17], and based on these goals, the HCP could then provide weekly asynchronous digital coaching individualized for each participant. The HCP would inspire, commend goal attainment, and motivate the participant. Furthermore, the participants could register their own self-reported weight measured at home. Liva is built with the option to record and track individual weight every day, providing multiple measuring points. The Liva app also has an option to track data collected via Apple and Google Fitbit, as well as all other devices connected via Validic. The primary data that are imported are step data and daily activity. There were no specific requirements for their home measurements regarding calibration, type, etc. The participants were advised to always use the same scale to weigh themselves and were instructed that they should preferably do it on the same day of the week, for example, Sunday morning without clothes on (with underwear) but without shoes and after they had been

to the toilet. This ensures the most uniform weight registration possible. The participants had to manually register their weights. Now, Liva offers synchronized bathing scales via an app so that data on weight, body composition, fat percentage, etc are measured and recorded automatically. But unfortunately, that was not a possibility when the Liva study was conducted and therefore, such parameters were not included. The program is also set up so that you receive notifications on your goals. If a participant has not registered a weight measurement on a certain day, he/she will receive a reminder. The mHealth tool is described in further detail in the Template of the Intervention Description and Replication [16]. As described earlier, the clinical weight measurements were taken at 6 and 12 months of clinical follow-up. To examine the agreement between these 2 measurement methods, we first had to define the limits for which self-reported weights could be used for the statistical analyses. As weight can change relatively quickly, the duration between the 2 measurements had to be reasonably close. To be included as a valid self-reported weight, the data point had to be 1-21 days prior to the 6 and 12 months of clinical follow-up. To minimize bias, we excluded self-reported weights on the same day or right after the clinical follow-up since our data showed that these self-reported weights were identical to the clinical measured weights (similar all the way down to decimals). This resulted in a total of 104 participants having a valid home measurement 1-21 days prior to the clinical assessments. The participants in this study did not know their clinical weights before registering their self-reported weights. [Figure 1](#) shows screenshots of the Liva Healthcare app.

**Figure 1.** Screenshots of the Liva Healthcare app showing certain features, including weight measurement, tracking, and weight goals.



## HCPs in This Study

The digital lifestyle coaching was provided by an HCP through the mHealth tool. All the HCPs were educated as nurses, dietitians, physiotherapists, or occupational therapists. They all received special training on how to practice digital health coaching and had practiced it for at least 2 years. One primary HCP was assigned to each participant to achieve and secure a close and trusting professional relationship [16].

## Statistical Analysis

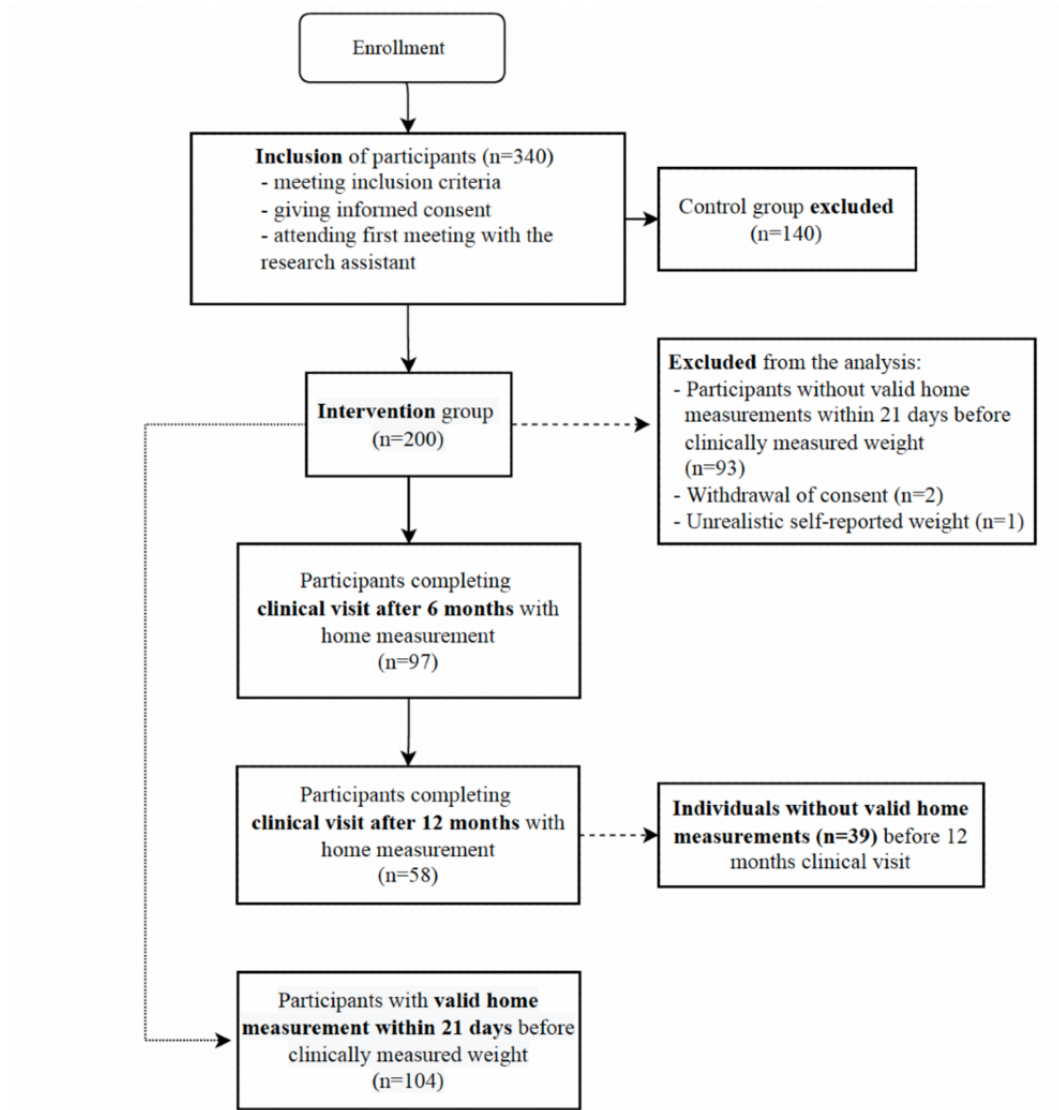
Differences in baseline characteristics between participants with and without valid home measurements were compared with analysis of variance for the continuous variables and chi-square test for the categorical variables. The following factors were included: gender, age, diabetes (yes/no), education, marital status, occupational status, baseline weight, and baseline BMI. Measured and self-reported weights were compared by linear regression (95% CI), and agreement was evaluated by Pearson correlation coefficients to determine the strength of the linear relationship. The degree of agreement between the self-reported and measured weight was also evaluated visually using Bland-Altman plots, and 95% limits of agreement were reported [18]. To identify determinants associated with the difference between measured and self-reported weights, we used linear regression and two-sided *t* test. Differences (clinical weight – self-reported weight) indicate if self-report was under (+) or over (–) estimated. Follow-up clinically measured weight was also used to determine the amount of weight change in participants classified as either achieving or not achieving the goal of 5% weight loss, which according to research is defined as a clinically relevant weight loss [19]. Mean values with corresponding standard deviations (SD) and frequencies with percentages have been reported. All analyses were performed using Stata version 13 (StataCorp LLC).

## Results

### Participant Characteristics

This study consisted of 200 participants from the intervention group of an RCT. Data were available after 6 months and 12 months, but 93 participants did not have a valid home measurement 1-21 days prior to their clinical weight measurement and were therefore excluded from the final analyses. Further, 3 participants were excluded because of withdrawal of consent and an unrealistic self-reported weight, with a 42-kg difference. As presented in [Figure 2](#), the final sample consisted of 104 participants with a valid home measurement, of which 97 and 58 participants were present at the 6-month and 12-month follow-up, respectively. There were no demographic differences at baseline when divided into groups with and without a valid home measurement. Participants' mean body weight was 103.9 kg, mean BMI was 35.3 kg/m<sup>2</sup>, and mean age was 52.1 years ([Table 1](#)). At 6 months and 12 months, 46 (44.2%) and 7 (6.7%) participants of the total 104 participants only had 1 self-reported weight, respectively, while 51 (49.1%) participants self-reported weights after both 6 and 12 months. Baseline characteristics of the participants with either 1 or 2 valid home measurements at 6-month and 12-month follow-ups did not differ in prevalence besides marital status. No differences in age, sex, disease, education, occupational status, and body composition were found ([Multimedia Appendix 1](#)). Furthermore, weight loss at 6 and 12 months did not differ between participants with and without a valid home measurement (data not shown).

[Multimedia Appendix 2](#) shows the percentage distribution of the days between clinically measured weight and self-reported weights of the 104 participants within 1-21 days. At 6 months, 60 (57.6%) and 78 (75%) weight registrations were made within 7 and 13 days, respectively. At 12 months, 64 (61.5%) and 88 (84.6%) weight registrations were made within 7 and 13 days.

**Figure 2.** Participant flow during the study.



**Table 1.** Baseline characteristics of the participants in the study group without and with home measurements.

	Without home measurement	With home measurement	All	P value
Participants, n (%)	96 (47.2)	104 (52.8)	200 (100)	N/A <sup>a</sup>
Age (years), mean (SD)	51.8 (11.3)	52.4 (9.4)	52.1 (10.3)	.69
<b>Sex, n (%)</b>				.31
Female	57 (61.3)	71 (68.3)	128 (65)	
Male	36 (38.7)	33 (31.7)	69 (35)	
<b>Diabetes, n (%)</b>				.43
Yes	49 (52.7)	49 (47.1)	98 (49.7)	
No	44 (47.3)	55 (52.9)	99 (50.3)	
<b>Education, n (%)</b>				.96
None	14 (15.1)	14 (13.5)	28 (14.2)	
Short (vocational courses, not university level)	23 (24.7)	26 (25)	49 (24.9)	
Long (university level, bachelors and masters)	9 (9.7)	10 (9.6)	19 (9.6)	
Middle (university level, bachelors)	45 (48.4)	53 (51)	98 (49.7)	
Don't know	2 (2.2)	1 (1)	3 (1.5)	
<b>Marital status, n (%)</b>				.17
Married	55 (59.1)	77 (74)	132 (67)	
Unmarried	23 (24.7)	16 (15.4)	39 (19.8)	
Divorced	13 (14)	10 (9.6)	23 (11.7)	
Widowed	2 (2.2)	1 (1)	3 (1.5)	
<b>Occupational status, n (%)</b>				.13
Employed	62 (66.7)	79 (76)	141 (71.6)	
Out of work (including on maternity leave or unemployment benefits)	8 (8.6)	8 (7.7)	16 (8.1)	
Out of work (social benefits)	5 (5.4)	0 (0)	5 (2.5)	
Early retirement	5 (5.4)	2 (1.9)	7 (3.6)	
Retired	11 (11.8)	14 (13.5)	25 (12.7)	
Student	2 (2.2)	1 (1)	3 (1.5)	
Weight (kg), mean (SD)	104.9 (17)	102.9 (14.3)	103.9 (15.6)	.36
BMI, mean (SD)	35.7 (4)	34.9 (3.6)	35.3 (3.8)	.14

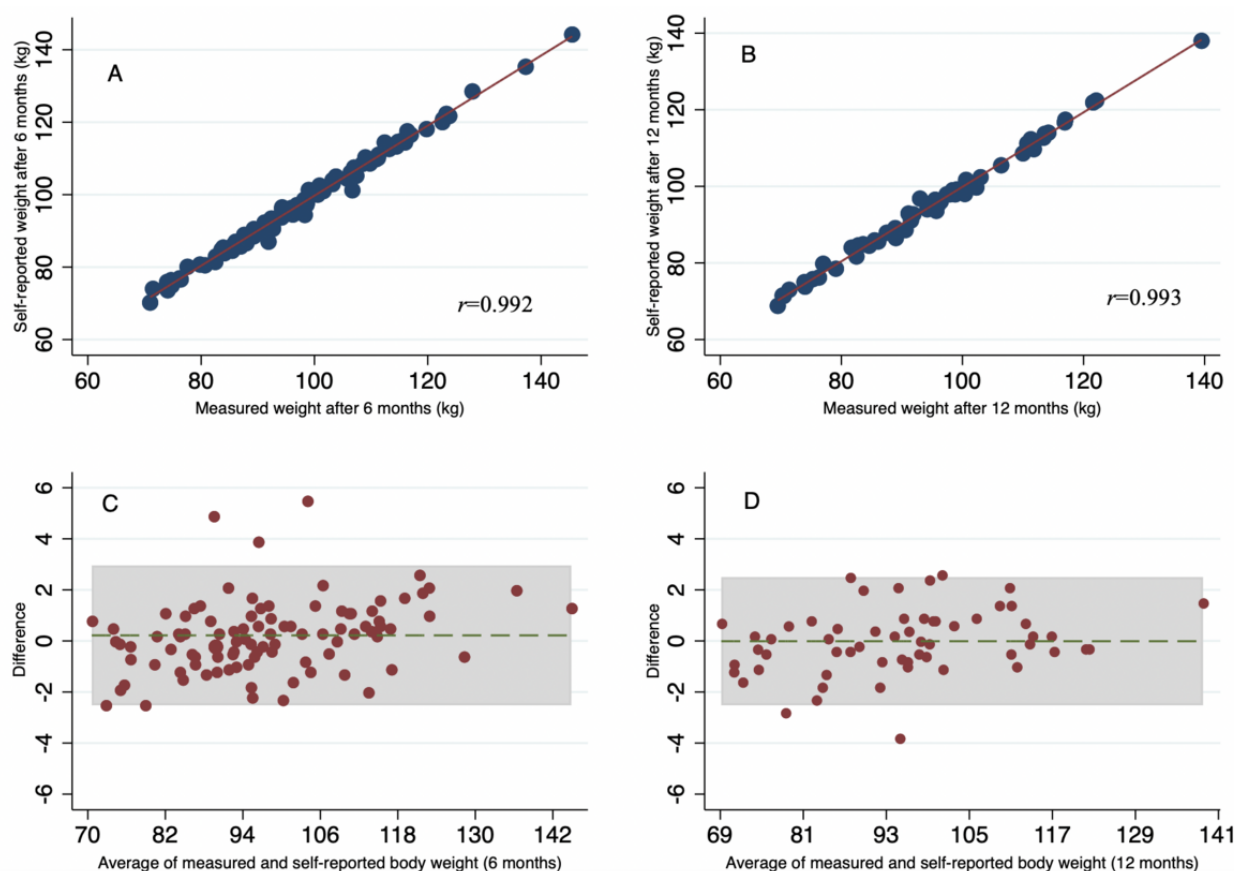
<sup>a</sup>N/A: not applicable.

### Overall Difference Between Clinically Measured Weight and Self-reported Body Weight

The average difference between measured and self-reported body weights at the 6-month follow-up was 1.03 kg (95% CI 1.01-1.05;  $P < .001$ ). The average difference between measured and self-reported body weights at the 12-month follow-up was

also 1.03 kg (95% CI 0.99-1.04;  $P < .001$ ). The Pearson correlation coefficient between measured and self-reported body weights showed a high correlation after 6 months ( $r = 0.99$ ) and 12 months of follow-up ( $r = 0.99$ ). The Bland-Altman plot showed a tendency of increased underestimation with greater clinically measured weight values (Figure 3).

**Figure 3.** (A-B) Linear regression (95% CI) of measured and self-reported weights after 6 and 12 months of clinical follow-up. (C-D) Bland-Altman plot of difference between measured and self-reported weights (y-axis) in relation to the average of measured weight (x-axis) after 6 and 12 months of clinical follow-up. The solid area shows mean difference (2 SD) and the dashed line shows differences equal to zero. A negative sign in difference indicates overestimation. A positive sign indicates underestimation of self-reported weight.

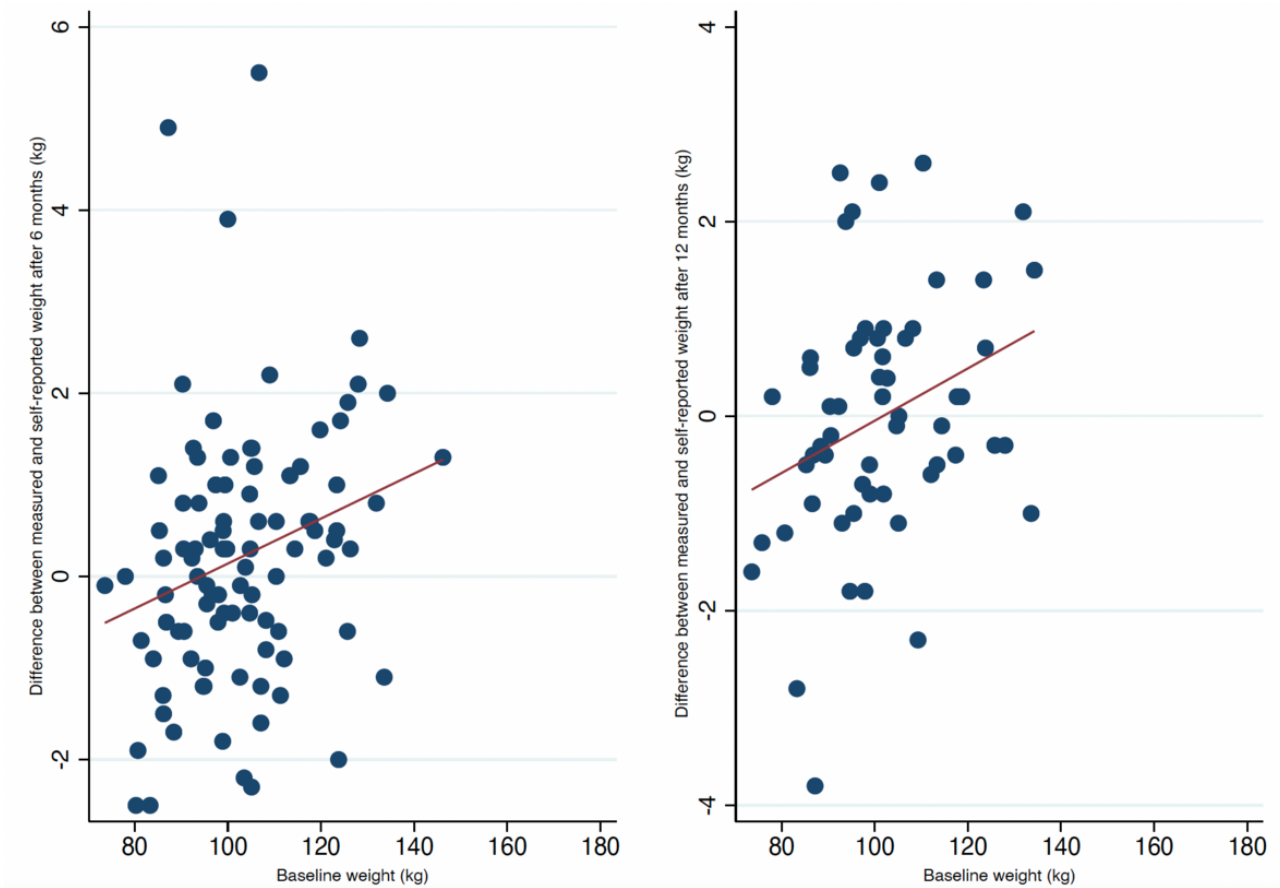


### Possible Discrepancies and Predicting the Extent of Misreporting Among Participants

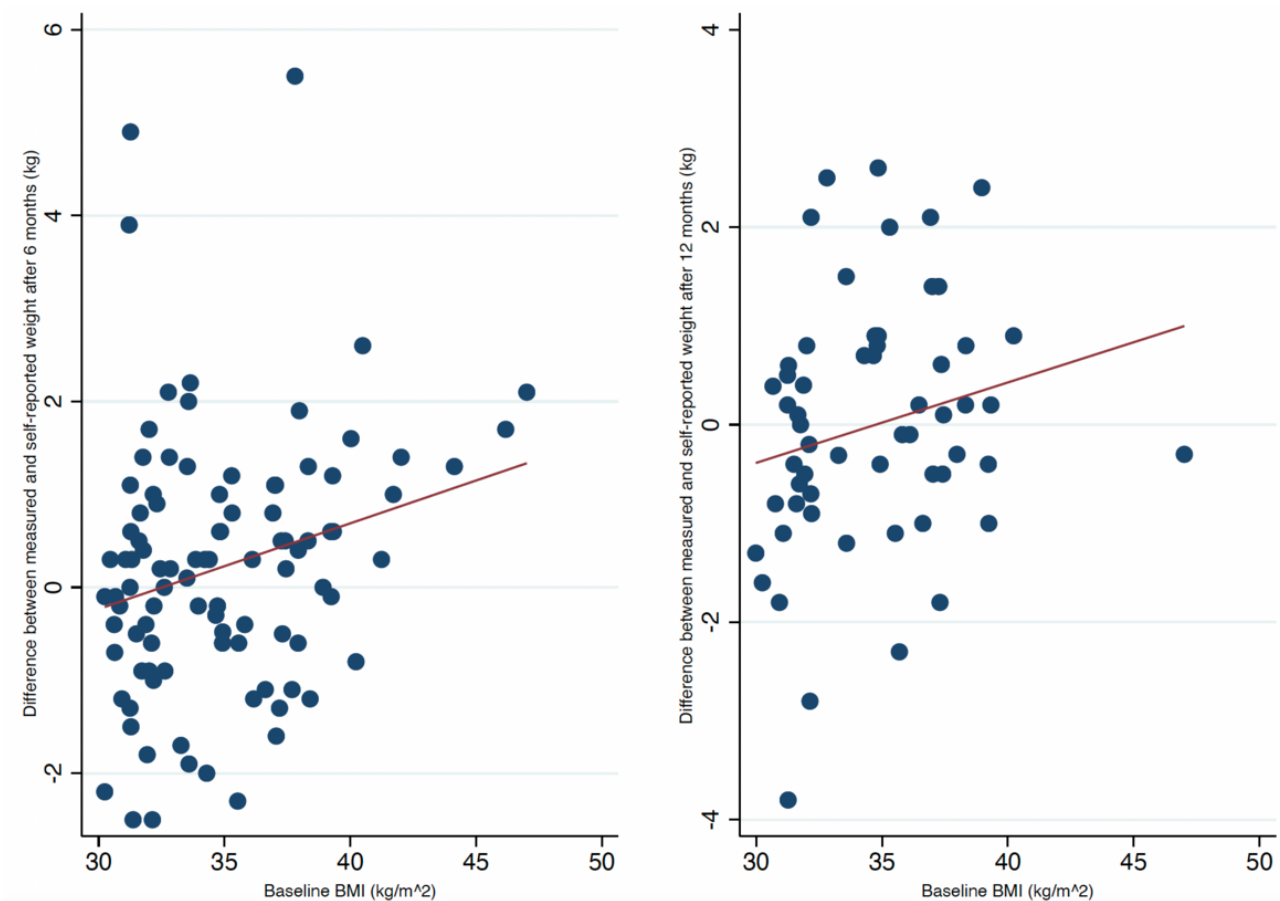
Baseline weight and BMI were associated with a discrepancy of 0.03 kg (95% CI 0.01-0.04;  $P=.01$ ) and 0.09 kg (95% CI 0.02-0.17;  $P=.02$ ) per increment of 1 kg and 1 kg/m<sup>2</sup>, respectively, between measured and self-reported weights (Figure 4 and Figure 5). Furthermore, weight change at 6 months was also associated with a difference of 0.1 kg (95% CI 0.04-0.01;  $P<.001$ ) per kilogram of weight change between measured and self-reported weights (Figure 6). Achievement of the 5% weight loss goal was associated with a difference of -0.28 kg (95% CI -0.59 to -0.03) at 6 months. Those who did not achieve the 5% weight loss had a difference of 0.79 kg (95% CI 0.34-1.23), with a between-group difference of 1.08 kg (95% CI 0.54-1.60;  $P<.001$ ) (Table 2). A within-group analysis was

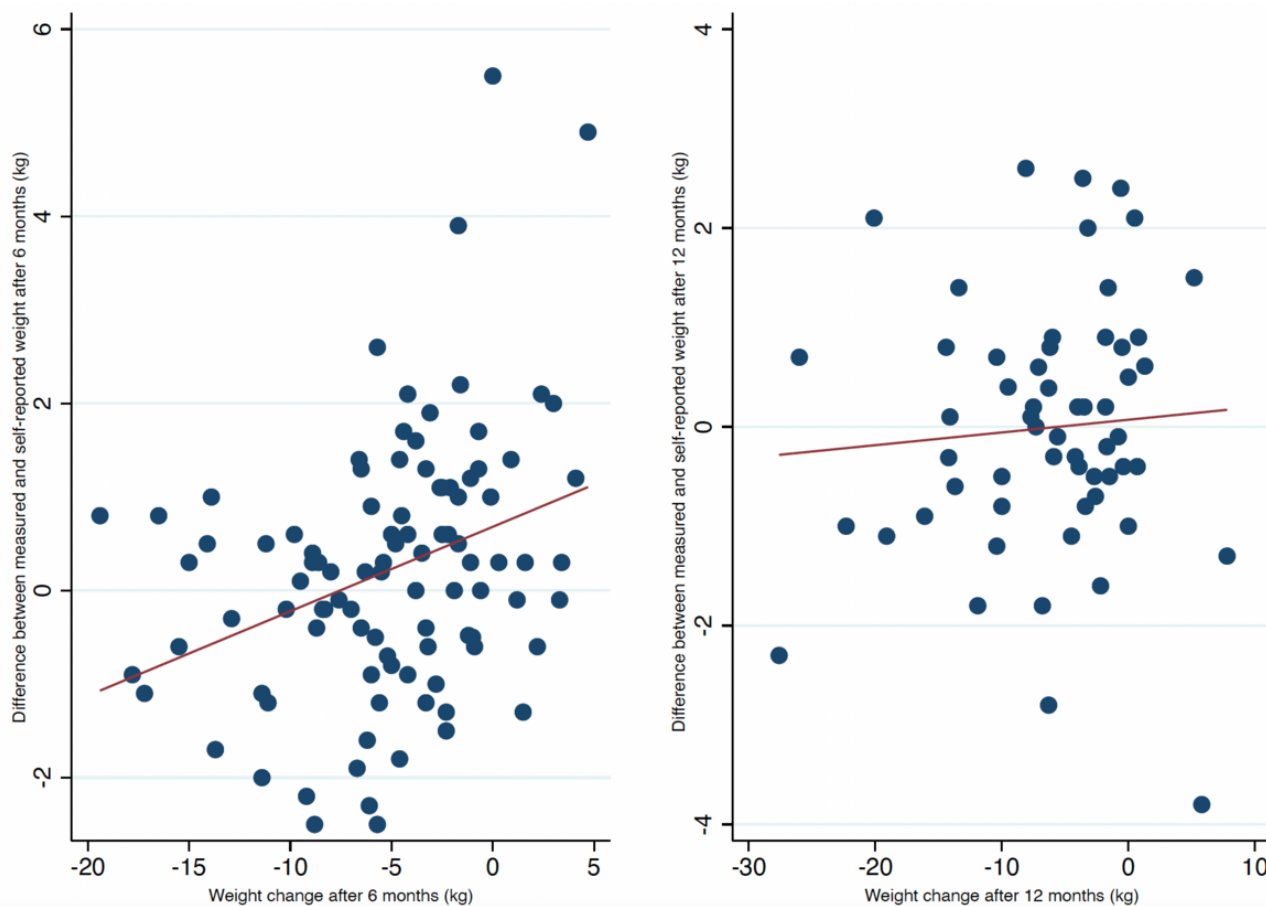
performed, and there were no significant differences between measured and self-reported weights when grouped by achievement/nonachievement of the 5% weight loss goal (Table 3). After 12 months, only baseline weight was associated with a discrepancy of 0.03 kg (95% CI 0.01-0.05;  $P=.02$ ) per increment of kilogram between measured and self-reported body weights (Figure 4). Baseline BMI, weight change, and achievement of the 5% weight loss goal were not associated with discrepancies after 12 months. Educational status, marital status, employment status, and days between weight measurements were not associated with the differences in clinically measured and self-reported weights (not shown). Fewer participants self-reported their weight prior to the 12-month follow-up (Figure 2), but as shown in Multimedia Appendix 1, there were no significant differences from the baseline values of those who self-reported twice and once.

**Figure 4.** Scatter plots with fitted lines at 6 and 12 months of clinical follow-up grouped by baseline weight.



**Figure 5.** Scatter plots with fitted lines at 6 and 12 months of clinical follow-up grouped by baseline BMI.



**Figure 6.** Scatter plots with fitted lines at 6 and 12 months of clinical follow-up grouped by weight changes during the intervention.**Table 2.** Predicting discrepancy between clinically measured and self-reported weights by 5% weight loss goal attainment at 6 and 12 months of clinical follow-up (between-group difference).

Between-group difference	At 6 months (n=97)				At 12 months (n=58)			
	n	Mean (SE)	95% CI	P value	n	Mean (SE)	95% CI	P value
Achievement of 5% weight loss goal	56	-0.28 (0.15)	-0.59 to -0.03		32	-0.15 (0.20)	-0.58 to 0.26	
Did not achieve 5% weight loss goal	41	0.79 (0.22)	0.34 to 1.23		26	0.17 (0.27)	-0.39 to 0.73	
Difference		1.06 (0.26)	0.54 to 1.60	<.001		0.32 (0.34)	-0.35 to 1.00	.34

**Table 3.** Predicting discrepancy between clinically measured and self-reported weights by 5% weight loss goal attainment at 6 and 12 months of clinical follow-up (within-group difference).

Within-group difference	At 6 months (n=97)				At 12 months (n=58)			
	n	Mean (SE)	95% CI	P value	n	Mean (SE)	95% CI	P value
Achievement of 5% weight loss goal	56	0.22 (0.19)	-0.16 to 0.59	.25	32	0.03 (0.23)	-0.44 to 0.50	.90
Did not achieve 5% weight loss goal	41	0.22 (0.22)	-0.22 to 0.67	.33	26	-0.06 (0.25)	-0.57 to 0.45	.80

## Discussion

### Principal Findings

Our primary analysis showed a small, albeit statistically significant, average difference of 1.03 kg ( $P < .001$ ) between measured and self-reported weights both after 6 and 12 months. The linear relationship between the clinically measured and self-reported weights was strongly correlated with high levels of agreement. Although the results show a significant difference,

the magnitude of these differences was quite small, suggesting negligible clinical importance. Fluctuations in weight vary in the range of 0.5-1 kg per day [20-22], and a typical bathroom scale has an uncertainty of 1%-2% [23]. This supports the clinical validity of self-reported weight despite the modest discrepancy of 1.03 kg in our study. Moreover, participants maintained their reports over 12 months. Thus, self-reported mobile-based weights may be adequate and reliable for monitoring weight change during a coaching program. When grouped by different demographic and lifestyle factors, we found

from our secondary analysis that participants with a higher baseline weight and BMI tended to underestimate their own weight by 0.03 kg and 0.09 kg at the 6-month follow-up, respectively. Furthermore, those who gained weight during the intervention also tended to underestimate their weight by 0.1 kg at the 6-month follow-up. Participants who achieved the 5% weight loss goal overestimated their weight by 0.28 kg. However, participants who did not achieve the 5% weight loss goal underestimated their weight by 0.79 kg at the 6-month follow-up. Although the within-group analysis did not show any differences, the between-group analysis indicated that those who achieved a 5% weight loss self-report more in accordance with their clinical weight. The discrepancies shown here were small, suggesting limited clinical relevance. None of the other demographic factors showed any significant discrepancies. Interestingly, the discrepancies improved over time when we analyzed the data from 6 to 12 months. As seen in the scatter plots, the data change from moving in a linear to a more constant pattern (closer to 0 in difference), which indicates that the previous discrepancy from baseline BMI and especially weight change improved from 6 to 12 months of clinical follow-up. This is also seen in the within-group analysis (Table 2), with an absolute change in difference of 0.22 kg to 0.03 kg in the achievement group and 0.22 kg to -0.06 kg in the nonachievement group, when we compare from 6 to 12 months (not significant). Furthermore, the between-group difference also improved by 1.06 kg to 0.32 kg from 6 to 12 months.

### Comparisons With Prior Research

Only few studies [13-15] have examined the agreement between measured and self-reported weights, with all of them being either paper-based or web-based through a web-based survey. To our knowledge, this is the first study to examine the agreement between clinically measured and self-reported weights from an mHealth-based lifestyle coaching program over 12 months. Our results show that the mobile-based reporting of own weight is a satisfactory method of data collection, which has also been proven in several international studies with web-based data collection [13-15,24]. Ekström et al [14] validated self-reported height, weight, and BMI among Swedish adolescents aged approximately 16 years by using a web-based survey. They found a mean difference of 1.1 kg between measured and self-reported weights, which was approximately the same as that found in our study (1.03 kg). Harvey-Berino et al [24] examined the agreement between measured and self-reported weights in a 6-month web-based obesity program. They found a mean underestimation of 0.86 kg. This overall positive agreement between self-reported and measured weight was further established in a nationwide cohort of 2643 US adults, which also found a relatively small underestimation [25]. According to our findings, baseline weight and BMI showed significant discrepancy, which agrees with several studies examining different populations [13-15,25]. Neermark et al [26] even found that calibrated values of self-reported BMI improved the predictive value of BMI for the risk of diabetes. Furthermore, a key finding from our study is that agreement between measured and self-reported weights appeared to worsen when participants gained weight and vice versa. Participants who were successful in losing weight between follow-ups reported

a more accurate weight. Only few studies have examined weight change, but those that did, found the same results [24,27,28].

We found no other discrepancies in our data when participants were grouped by different demographic factors. This contrasts with other studies, which suggest that women tend to underestimate their own weight [13-15,25]. According to these studies, a possible explanation for this could be that women tend to be more aware of their own weight because of societal and psychological factors shaping their body image views. However, we found no statistical differences by gender. Koebnick et al [28] examined factors related to depressive symptoms among 17-year-old girls and found that strong experience of negative emotions such as anger, anxiety, and contempt was associated with underestimation of body weight. Furthermore, lower body satisfaction was associated with higher BMI, which led to higher negative emotions. Another key finding from our data indicates that a 12-month coaching intervention improves the discrepancy between measured and self-reported weights. Possible explanations for this could be that (1) the participants are regularly being observed and motivated by a coach with whom they have a good relationship between follow-ups, (2) they must measure their own weight at follow-up anyway, and (3) weight could be more stable after 12 months and thus, the difference of weight measured 21 days before the clinical visit is smaller. However, we cannot say whether this correlation implies causation. Our study design could not test whether being part of a coaching program improves the agreement between self-reported and clinically measured weights. In an optimal setting, there should have been a control group who did not receive coaching but still had access to the app and registered their weight. A previous study [29] demonstrated that regular feedback improved the validity of self-reported weight among obese employees. Contrary to the finding in our study, Jerome et al [27] found that the magnitude of underestimation doubled between 6 and 24 months of clinical follow-up. However, they also found that weight loss was associated with higher validity, which agrees with our and other studies. Jerome et al [27] also found that those with self-reported weight lost 3 times more weight compared to those without self-reported weight. This is important, because 39 (37.5%) of our 104 participants did not have a self-reported weight prior to the 12-month clinical follow-up. It can be assumed that lack of weight loss could have contributed to lesser motivation in this small group, resulting in fewer weight registrations. However, we did not find any differences regarding weight change between participants with and without self-reported weight. It is not quite clear whether weight loss encouraged accurate self-assessments or vice versa. Nevertheless, digital weight loss programs should be aware of this tendency.

Several differences in our study design might explain the different results. In our study, participants were older, more overweight, and had T2D, which could decrease the validity compared to populations with fewer overweight participants and with no chronic illnesses. Age could also impact the validity, because studies [30,31] show that younger populations have a higher usage of health apps and therefore are more prone to higher user engagement. Furthermore, we only chose self-reported weights 1-21 days prior to each clinical follow-up,

with 65.3% (68/104) and 75% (78/104) of the participants self-reporting their weights within 7 and 13 days, respectively ([Multimedia Appendix 2](#)). This contrasts with that reported in other studies [14,15] where the limit of the self-reported measurements is further away from the date of the clinical measurements, although one study [27] had a limit of minimum 7 days. However, our data show that the number of days between measuring weights was not associated with discrepancy, including self-reported weights within 10 days. Jerome et al [27] also included self-reported weights from the same day as the clinical follow-up. We chose to exclude the self-reported weights that were made on the same day or right after the clinical measurements, since our data showed that many of the self-reported weights were identical to the clinically measured weights (similar all the way down to decimals). All the participants in our study were overweight or had T2D and were motivated to lose weight. This may limit the generalizability, since validity can vary with age, ethnicity, diseases, weight loss motivation, and several other variables. It is important to carry out validation studies in different countries and populations before generalization.

### Strengths and Limitations

The main strength of this study is the duration of the follow-up (12 months) compared to that in other studies, which made it possible to examine the agreement over time and assess whether a digital lifestyle coaching intervention had an impact on the validity. Our study sample was relatively large, taking into consideration that it was primary care-anchored and all the participants were overweight with T2D and motivated to lose weight. The participants were also not aware of this study, which

could minimize potential bias. Clinical weight measurements were made according to a protocol by HCPs with a fair limit for days between self-reported and clinically measured weight.

The limitations in our study are as follows. In our study, first, only 58 (55.7%) out of 104 participants measured and self-reported weights at 12-month follow-up, which could induce a nonresponse bias and therefore limit the generalizability of our findings and affect the statistical power. However, baseline characteristics showed no differences ([Multimedia Appendix 1](#)), and weight change/loss was not significantly different between participants with and without self-reported weights. Second, we must also consider that weight fluctuations can happen during the day or during the menstrual cycle for females. However, from our analysis, gender did not influence any discrepancies. Lastly, clinical weight was measured with light clothes. Furthermore, the use of different measuring equipment by participants at home may have introduced measurement bias. We did not find any large systematic differences in the agreement between the clinically measured and self-reported weights.

### Conclusion

In this mHealth-based clinical trial, we found a high level of agreement between self-reported and clinically measured weights. Self-reported weight was, on average, underestimated by 1.03 kg. Baseline weight, baseline BMI, and weight change influenced this discrepancy. Our findings suggest that self-reported weights from an mHealth-based solution is a valid method for anthropometric measurements in a digital lifestyle intervention. Future studies should include longer follow-up periods with repeated measures over time.

---

### Acknowledgments

CJB initiated this study in continuation of research in his PhD dissertation, which was done at the Research Unit for General Practice at University of Southern Denmark (thesis defended on August 2018) [16]. This study was partly funded by Liva Healthcare A/S. A formal research agreement has been made between Liva Healthcare A/S and the University of Southern Denmark to guide the running and financial aspects of the project. The patients in this study were not economically compensated for their participation. Lastly, we want to thank TBO for assisting us during the statistical analysis.

---

### Conflicts of Interest

CJB owns stock in Liva Healthcare A/S and originally cofounded Liva Healthcare A/S, the company that developed parts of the technical platform. CJB works currently at the Research Unit for General Practice at University of Southern Denmark. DHL is employed at Liva Healthcare A/S. DHL, TBO, JS, and AI have no financial interests in Liva Healthcare A/S or any other aspects of this study.

---

#### Multimedia Appendix 1

Baseline characteristics of the participants with either 1 or 2 valid home measurements.  
[\[PDF File \(Adobe PDF File\), 62 KB - formative\\_v6i9e40739\\_app1.pdf\]](#)

---

#### Multimedia Appendix 2

Percentage distribution of days between weights at 6 and 12 months of clinical follow-up.  
[\[PNG File, 59 KB - formative\\_v6i9e40739\\_app2.png\]](#)

---

### References

1. Global status report on noncommunicable diseases. In: World Health Organization . Geneva: World Health Organization ; 2014.

2. Afshin A, Babalola D, Mclean M, Yu Z, Ma W, Chen C, et al. Information Technology and Lifestyle: A Systematic Evaluation of Internet and Mobile Interventions for Improving Diet, Physical Activity, Obesity, Tobacco, and Alcohol Use. *J Am Heart Assoc* 2016 Aug 31;5(9):e003058 [FREE Full text] [doi: [10.1161/JAHA.115.003058](https://doi.org/10.1161/JAHA.115.003058)] [Medline: [27581172](https://pubmed.ncbi.nlm.nih.gov/27581172/)]
3. Ried-Larsen M, Christensen R, Hansen KB, Johansen MY, Pedersen M, Zacho M, et al. Head-to-head comparison of intensive lifestyle intervention (U-TURN) versus conventional multifactorial care in patients with type 2 diabetes: protocol and rationale for an assessor-blinded, parallel group and randomised trial. *BMJ Open* 2015 Dec 09;5(12):e009764 [FREE Full text] [doi: [10.1136/bmjopen-2015-009764](https://doi.org/10.1136/bmjopen-2015-009764)] [Medline: [26656025](https://pubmed.ncbi.nlm.nih.gov/26656025/)]
4. Johansen MY, MacDonald CS, Hansen KB, Karstoft K, Christensen R, Pedersen M, et al. Effect of an Intensive Lifestyle Intervention on Glycemic Control in Patients With Type 2 Diabetes: A Randomized Clinical Trial. *JAMA* 2017 Aug 15;318(7):637-646 [FREE Full text] [doi: [10.1001/jama.2017.10169](https://doi.org/10.1001/jama.2017.10169)] [Medline: [28810024](https://pubmed.ncbi.nlm.nih.gov/28810024/)]
5. Diabetes Prevention Program (DPP) Research Group, Hamman RF, Horton E, Barrett-Connor E, Bray GA, Christophi CA, et al. Factors affecting the decline in incidence of diabetes in the Diabetes Prevention Program Outcomes Study (DPPOS). *Diabetes* 2015 Mar;64(3):989-998 [FREE Full text] [doi: [10.2337/db14-0333](https://doi.org/10.2337/db14-0333)] [Medline: [25277389](https://pubmed.ncbi.nlm.nih.gov/25277389/)]
6. Haste A, Adamson AJ, McColl E, Araujo-Soares V, Bell R. Web-Based Weight Loss Intervention for Men With Type 2 Diabetes: Pilot Randomized Controlled Trial. *JMIR Diabetes* 2017 Jul 07;2(2):e14 [FREE Full text] [doi: [10.2196/diabetes.7430](https://doi.org/10.2196/diabetes.7430)] [Medline: [30291100](https://pubmed.ncbi.nlm.nih.gov/30291100/)]
7. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The Lancet* 2012 Jul;380(9838):219-229. [doi: [10.1016/S0140-6736\(12\)61031-9](https://doi.org/10.1016/S0140-6736(12)61031-9)]
8. Johansen MY, MacDonald CS, Hansen KB, Karstoft K, Christensen R, Pedersen M, et al. Effect of an Intensive Lifestyle Intervention on Glycemic Control in Patients With Type 2 Diabetes: A Randomized Clinical Trial. *JAMA* 2017 Aug 15;318(7):637-646 [FREE Full text] [doi: [10.1001/jama.2017.10169](https://doi.org/10.1001/jama.2017.10169)] [Medline: [28810024](https://pubmed.ncbi.nlm.nih.gov/28810024/)]
9. Lean ME, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *Lancet* 2018 Feb 10;391(10120):541-551. [doi: [10.1016/S0140-6736\(17\)33102-1](https://doi.org/10.1016/S0140-6736(17)33102-1)] [Medline: [29221645](https://pubmed.ncbi.nlm.nih.gov/29221645/)]
10. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. *J Am Diet Assoc* 2011 Jan;111(1):92-102 [FREE Full text] [doi: [10.1016/j.jada.2010.10.008](https://doi.org/10.1016/j.jada.2010.10.008)] [Medline: [21185970](https://pubmed.ncbi.nlm.nih.gov/21185970/)]
11. Tufano JT, Karras BT. Mobile eHealth interventions for obesity: a timely opportunity to leverage convergence trends. *J Med Internet Res* 2005 Dec 20;7(5):e58 [FREE Full text] [doi: [10.2196/jmir.7.5.e58](https://doi.org/10.2196/jmir.7.5.e58)] [Medline: [16403722](https://pubmed.ncbi.nlm.nih.gov/16403722/)]
12. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. *Obes Rev* 2010 Apr;11(4):306-321. [doi: [10.1111/j.1467-789X.2009.00646.x](https://doi.org/10.1111/j.1467-789X.2009.00646.x)] [Medline: [19754633](https://pubmed.ncbi.nlm.nih.gov/19754633/)]
13. Pursey K, Burrows TL, Stanwell P, Collins CE. How accurate is web-based self-reported height, weight, and body mass index in young adults? *J Med Internet Res* 2014 Jan 07;16(1):e4 [FREE Full text] [doi: [10.2196/jmir.2909](https://doi.org/10.2196/jmir.2909)] [Medline: [24398335](https://pubmed.ncbi.nlm.nih.gov/24398335/)]
14. Ekström S, Kull I, Nilsson S, Bergström A. Web-based self-reported height, weight, and body mass index among Swedish adolescents: a validation study. *J Med Internet Res* 2015 Mar 18;17(3):e73 [FREE Full text] [doi: [10.2196/jmir.3947](https://doi.org/10.2196/jmir.3947)] [Medline: [25791395](https://pubmed.ncbi.nlm.nih.gov/25791395/)]
15. Chai LK, Collins CE, May C, Holder C, Burrows TL. Accuracy of Parent-Reported Child Height and Weight and Calculated Body Mass Index Compared With Objectively Measured Anthropometrics: Secondary Analysis of a Randomized Controlled Trial. *J Med Internet Res* 2019 Sep 16;21(9):e12532 [FREE Full text] [doi: [10.2196/12532](https://doi.org/10.2196/12532)] [Medline: [31538954](https://pubmed.ncbi.nlm.nih.gov/31538954/)]
16. Brandt CJ, Christensen JR, Lauridsen JT, Nielsen JB, Søndergaard J, Sortsø C. Evaluation of the Clinical and Economic Effects of a Primary Care Anchored, Collaborative, Electronic Health Lifestyle Coaching Program in Denmark: Protocol for a Two-Year Randomized Controlled Trial. *JMIR Res Protoc* 2020 Jun 25;9(6):e19172 [FREE Full text] [doi: [10.2196/19172](https://doi.org/10.2196/19172)] [Medline: [32584260](https://pubmed.ncbi.nlm.nih.gov/32584260/)]
17. Ryan P. Integrated Theory of Health Behavior Change: background and intervention development. *Clin Nurse Spec* 2009;23(3):161-70; quiz 171 [FREE Full text] [doi: [10.1097/NUR.0b013e3181a42373](https://doi.org/10.1097/NUR.0b013e3181a42373)] [Medline: [19395894](https://pubmed.ncbi.nlm.nih.gov/19395894/)]
18. Flegal KM, Graubard B, Ioannidis JPA. Use and reporting of Bland-Altman analyses in studies of self-reported versus measured weight and height. *Int J Obes (Lond)* 2020 Jun;44(6):1311-1318 [FREE Full text] [doi: [10.1038/s41366-019-0499-5](https://doi.org/10.1038/s41366-019-0499-5)] [Medline: [31792334](https://pubmed.ncbi.nlm.nih.gov/31792334/)]
19. Ryan DH, Yockey SR. Weight Loss and Improvement in Comorbidity: Differences at 5%, 10%, 15%, and Over. *Curr Obes Rep* 2017 Jun;6(2):187-194 [FREE Full text] [doi: [10.1007/s13679-017-0262-y](https://doi.org/10.1007/s13679-017-0262-y)] [Medline: [28455679](https://pubmed.ncbi.nlm.nih.gov/28455679/)]
20. Ma Y, Olendzki BC, Li W, Hafner AR, Chiriboga D, Hebert JR, et al. Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population. *Eur J Clin Nutr* 2006 Apr;60(4):519-528 [FREE Full text] [doi: [10.1038/sj.ejcn.1602346](https://doi.org/10.1038/sj.ejcn.1602346)] [Medline: [16340952](https://pubmed.ncbi.nlm.nih.gov/16340952/)]
21. Orsama A, Mattila E, Ermes M, van Gils M, Wansink B, Korhonen I. Weight rhythms: weight increases during weekends and decreases during weekdays. *Obes Facts* 2014;7(1):36-47 [FREE Full text] [doi: [10.1159/000356147](https://doi.org/10.1159/000356147)] [Medline: [24504358](https://pubmed.ncbi.nlm.nih.gov/24504358/)]

22. Turicchi J, O'Driscoll R, Horgan G, Duarte C, Palmeira AL, Larsen SC, et al. Weekly, seasonal and holiday body weight fluctuation patterns among individuals engaged in a European multi-centre behavioural weight loss maintenance intervention. *PLoS One* 2020;15(4):e0232152 [FREE Full text] [doi: [10.1371/journal.pone.0232152](https://doi.org/10.1371/journal.pone.0232152)] [Medline: [32353079](https://pubmed.ncbi.nlm.nih.gov/32353079/)]
23. Yorkin M, Spaccarotella K, Martin-Biggers J, Quick V, Byrd-Bredbenner C. Accuracy and consistency of weights provided by home bathroom scales. *BMC Public Health* 2013 Dec 17;13:1194 [FREE Full text] [doi: [10.1186/1471-2458-13-1194](https://doi.org/10.1186/1471-2458-13-1194)] [Medline: [24341761](https://pubmed.ncbi.nlm.nih.gov/24341761/)]
24. Harvey-Berino J, Krukowski RA, Buzzell P, Ogden D, Skelly J, West DS. The accuracy of weight reported in a web-based obesity treatment program. *Telemed J E Health* 2011 Nov;17(9):696-699 [FREE Full text] [doi: [10.1089/tmj.2011.0032](https://doi.org/10.1089/tmj.2011.0032)] [Medline: [21882997](https://pubmed.ncbi.nlm.nih.gov/21882997/)]
25. Hodge JM, Shah R, McCullough ML, Gapstur SM, Patel AV. Validation of self-reported height and weight in a large, nationwide cohort of U.S. adults. *PLoS One* 2020;15(4):e0231229 [FREE Full text] [doi: [10.1371/journal.pone.0231229](https://doi.org/10.1371/journal.pone.0231229)] [Medline: [32282816](https://pubmed.ncbi.nlm.nih.gov/32282816/)]
26. Neermark S, Holst C, Bisgaard T, Bay-Nielsen M, Becker U, Tolstrup J. Validation and calibration of self-reported height and weight in the Danish Health Examination Survey. *Eur J Public Health* 2019;29(2):291-296. [doi: [10.1093/eurpub/cky187](https://doi.org/10.1093/eurpub/cky187)]
27. Jerome GJ, Dalcin A, Coughlin JW, Fitzpatrick S, Wang N, Durkin N, et al. Longitudinal accuracy of web-based self-reported weights: results from the Hopkins POWER Trial. *J Med Internet Res* 2014 Jul 15;16(7):e173 [FREE Full text] [doi: [10.2196/jmir.3332](https://doi.org/10.2196/jmir.3332)] [Medline: [25042773](https://pubmed.ncbi.nlm.nih.gov/25042773/)]
28. Koebnick C, Saksvig B, Li X, Sidell M, Wu TT, Young DR. The Accuracy of Self-Reported Body Weight Is High but Dependent on Recent Weight Change and Negative Affect in Teenage Girls. *Int J Environ Res Public Health* 2020 Nov 06;17(21):8203 [FREE Full text] [doi: [10.3390/ijerph17218203](https://doi.org/10.3390/ijerph17218203)] [Medline: [33172095](https://pubmed.ncbi.nlm.nih.gov/33172095/)]
29. Pronk NP, Crain AL, Vanwormer JJ, Martinson BC, Boucher JL, Cosentino DL. The Use of Telehealth Technology in Assessing the Accuracy of Self-Reported Weight and the Impact of a Daily Immediate-Feedback Intervention among Obese Employees. *Int J Telemed Appl* 2011;2011:909248 [FREE Full text] [doi: [10.1155/2011/909248](https://doi.org/10.1155/2011/909248)] [Medline: [21760782](https://pubmed.ncbi.nlm.nih.gov/21760782/)]
30. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use Medical Apps Among Older Adults in the Netherlands: Cross-Sectional Study. *J Med Internet Res* 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: [10.2196/18080](https://doi.org/10.2196/18080)] [Medline: [32624465](https://pubmed.ncbi.nlm.nih.gov/32624465/)]
31. Mustafa AS, Ali N, Dhillon JS, Alkawsy G, Baashar Y. User Engagement and Abandonment of mHealth: A Cross-Sectional Survey. *Healthcare* 2022 Jan 24;10(2):221. [doi: [10.3390/healthcare10020221](https://doi.org/10.3390/healthcare10020221)]

## Abbreviations

- DHI:** digital health intervention  
**HCP:** health care professional  
**Liva:** Lifestyle change InterVention and mHealth Application  
**mHealth:** mobile health  
**RCT:** randomized controlled trial  
**T2D:** type 2 diabetes

*Edited by M Focsa; submitted 03.07.22; peer-reviewed by S Graham, T Sathish, J Avis; comments to author 09.08.22; revised version received 24.08.22; accepted 29.08.22; published 14.09.22.*

*Please cite as:*

*Imeraj A, Olesen TB, Laursen DH, Søndergaard J, Brandt CJ*

*Agreement Between Clinically Measured Weight and Self-reported Weight Among Patients With Type 2 Diabetes Through an mHealth Lifestyle Coaching Program in Denmark: Secondary Analysis of a Randomized Controlled Trial*

*JMIR Form Res* 2022;6(9):e40739

URL: <https://formative.jmir.org/2022/9/e40739>

doi: [10.2196/40739](https://doi.org/10.2196/40739)

PMID: [36047606](https://pubmed.ncbi.nlm.nih.gov/36047606/)

©Albi Imeraj, Thomas Bastholm Olesen, Ditte Hjorth Laursen, Jens Søndergaard, Carl Joakim Brandt. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 14.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# The Usability of a Smartphone-Based Fall Risk Assessment App for Adult Wheelchair Users: Observational Study

Mikaela Frechette<sup>1,2</sup>, BSc, MSc, PhD; Jason Fanning<sup>3</sup>, BSc, MSc, PhD; Katherine Hsieh<sup>4</sup>, BSc, PhD; Laura Rice<sup>1,5,6</sup>, BSc, MPT, PhD; Jacob Sosnoff<sup>5,7</sup>, BSc, MSc, PhD

<sup>1</sup>Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign, Urbana, IL, United States

<sup>2</sup>Siebel Center for Design, University of Illinois at Urbana-Champaign, Champaign, IL, United States

<sup>3</sup>Department of Health and Exercise Science, Wake Forest University, Winston-Salem, NC, United States

<sup>4</sup>Department of Internal Medicine, Section on Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States

<sup>5</sup>Illinois Multiple Sclerosis Research Collaborative, University of Illinois at Urbana-Champaign, Urbana, IL, United States

<sup>6</sup>Center on Health, Aging, and Disability, University of Illinois at Urbana-Champaign, Urbana, IL, United States

<sup>7</sup>Department of Physical Therapy, Rehabilitation Science, and Athletic Training, University of Kansas Medical Center, Kansas City, KS, United States

**Corresponding Author:**

Jacob Sosnoff, BSc, MSc, PhD

Department of Physical Therapy, Rehabilitation Science, and Athletic Training

University of Kansas Medical Center

3901 Rainbow Blvd

Kansas City, KS, 66103

United States

Phone: 1 913 588 5235

Email: [jsosnoff@kumc.edu](mailto:jsosnoff@kumc.edu)

## Abstract

**Background:** Individuals who use wheelchairs and scooters rarely undergo fall risk screening. Mobile health technology is a possible avenue to provide fall risk assessment. The promise of this approach is dependent upon its usability.

**Objective:** We aimed to determine the usability of a fall risk mobile health app and identify key technology development insights for aging adults who use wheeled devices.

**Methods:** Two rounds (with 5 participants in each round) of usability testing utilizing an iterative design-evaluation process were performed. Participants completed use of the custom-designed fall risk app, Steady-Wheels. To quantify fall risk, the app led participants through 12 demographic questions and 3 progressively more challenging seated balance tasks. Once completed, participants shared insights on the app's usability through semistructured interviews and completion of the Systematic Usability Scale. Testing sessions were recorded and transcribed. Codes were identified within the transcriptions to create themes. Average Systematic Usability Scale scores were calculated for each round.

**Results:** The first round of testing yielded 2 main themes: ease of use and flexibility of design. Systematic Usability Scale scores ranged from 72.5 to 97.5 with a mean score of 84.5 (SD 11.4). After modifications were made, the second round of testing yielded 2 new themes: app layout and clarity of instruction. Systematic Usability Scale scores improved in the second iteration and ranged from 87.5 to 97.5 with a mean score of 91.9 (SD 4.3).

**Conclusions:** The mobile health app, Steady-Wheels, has excellent usability and the potential to provide adult wheeled device users with an easy-to-use, remote fall risk assessment tool. Characteristics that promoted usability were guided navigation, large text and radio buttons, clear and brief instructions accompanied by representative illustrations, and simple error recovery. Intuitive fall risk reporting was achieved through the presentation of a single number located on a color-coordinated continuum that delineated low, medium, and high risk.

(*JMIR Form Res* 2022;6(9):e32453) doi:[10.2196/32453](https://doi.org/10.2196/32453)

**KEYWORDS**

usability testing; mobile health; wheeled device user; fall risk; telehealth; mHealth; mobile device; smartphone; health applications; older adults; elderly population; device usability

## Introduction

Over 3 million individuals in the United States require the use of a wheelchair for mobility [1] and wheelchair use is expected to increase [2,3]. Although wheeled device use has numerous benefits [4], it presents several unique risks, such as falls. Roughly 75% of wheelchair users fall at least once a year [5-8]. Approximately 50% of reported falls cause injuries [7], ranging from minor (ie, abrasions) to serious (eg, fractures) [6]. Falls can also induce fear of falling [9] and activity curtailment [7], which are associated with isolation and decreased independence and quality of life [10].

Falls are detrimental to wheeled device users' health and well-being, making fall risk screening a necessary part of overall health care. Although the US Centers for Disease Control and Prevention recommends annual fall risk screening for older adults, current screening recommendations are designed for ambulatory adults [11]. Moreover, fall risk screening is rarely performed in clinical practice, and there are numerous barriers to the implementation of effective fall prevention programs for wheelchair users. As a result, most individuals who rely on wheeled mobility do not undergo routine fall risk screening. Additionally, the COVID-19 pandemic and its related restrictions necessitate remote monitoring of health. This highlights the need for novel remote fall risk technology specific to this population.

Due to limited access, researchers are exploring innovative approaches to deliver comprehensive and objective fall risk assessment to wheeled device users. One possible method leverages the capabilities of smartphone technology by developing an at-home fall risk health app [12-16]. This approach has been examined in ambulatory adults with a range of physical function [17,18]. Building on this potential, it has been demonstrated that a smartphone-based approach is a valid and reliable method to distinguish wheeled device users with and without impaired seated postural control [19]. Collectively, these findings provide the rationale for the development of an

objective mobile health app that can provide wheeled device users with at-home fall risk assessment.

Although there is a strong rationale for the development of this type of health app, ensuring that such a tool is easy to use and provides intuitive fall risk score reporting is a necessary precursor to its future use in health behavior interventions [20]. Consequently, the purpose of the current study is to determine the usability of a fall risk mobile health app, Steady-Wheels, and identify key technology development insights for aging adults who use wheeled devices. This health app is an adaptation of a pre-existing fall risk app for older adults [18]. Based on prior investigations, we hypothesized that this health app would have a high level of usability.

## Methods

### Underlying Design Considerations

When designing the first iteration of the health app, we considered our target users' (individuals aging with a physical disability) characteristics (Table 1). To ensure a high degree of usability, age-related changes and limitations due to disease or injury were taken into consideration, particularly as they related to cognitive overload, dexterity, and sensory function. To reduce cognitive overload, the app was designed to provide written instructions immediately preceding a task. Only one set of instructions was presented per slide, and large graphics depicting the task were also provided. This layout streamlined the app and reduced the need for working memory of the participants. In total, there were 14 slides, taking approximately 10 minutes to complete. Decrements in dexterity are commonly seen in those who have neurological complications [21-23] and age-associated arthritis [24]. To account for this within the app, selection options and buttons were made large, and typed responses were avoided. Sensory-related changes were accommodated by the use of black text written in a 14-point font on a white background [25], and auditory processing deficits were accommodated by the use of leading audio cues with simultaneous vibrations.

**Table 1.** Participant demographic information.

Characteristics	First iteration	Second iteration
Age (years), mean (SD)	59.0 (12.2)	58.0 (13.1)
<b>Sex, n (%)</b>		
Male	3 (60)	2 (40)
Female	2 (40)	3 (60)
Smartphone usage, n (%)	5 (100)	5 (100)
Time using mobility device (years), mean (SD)	25 (20.3)	25 (27.4)
<b>Primary mobility device, n (%)</b>		
Power chair	4 (80)	2 (40)
Manual chair	1 (20)	0
Scooter	0	3 (60)
<b>Reason for wheeled mobility, n (%)</b>		
Multiple sclerosis	2 (40)	4 (80)
Paraplegia/quadriplegia	2 (40)	1 (20)
Stroke	1 (20)	0
History of falls ( $\geq 1$ falls/year), n (%)	1 (20)	2 (40)
Self-reported fear of falling, n (%)	5 (100)	5 (100)
<b>Level of education, n (%)</b>		
High school graduate/General Educational Development Test Credential	0	1 (20)
Some or in-progress college/associate degree	2 (40)	0
Bachelor's degree	0	1 (20)
Master's degree	3 (60)	2 (40)
Doctoral degree	0	1 (20)

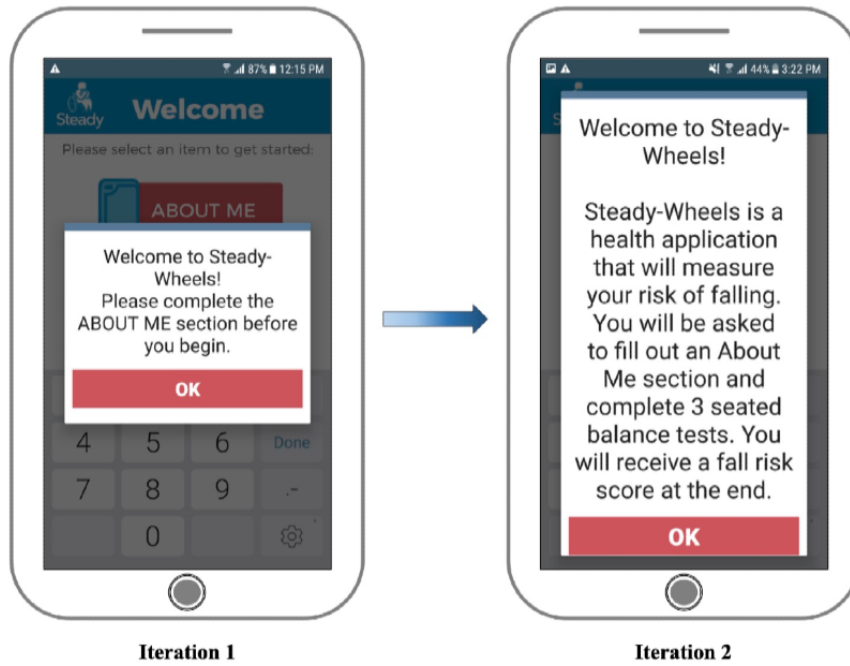
### Components of the Steady-Wheels App

The fall risk app, Steady-Wheels, was developed in Android Studio 3.1.2. Upon opening the app, users are presented with a welcome screen that outlines the purpose of the app and provides an overview of the process (Figure 1). Steady-Wheels has two main components: a patient-reported outcome section and a performance test section (Figure 2). The patient-reported outcome component asks the participant to complete a 13-item health history questionnaire (including age, sex, number of falls in the last year, and activities that provoke concerns about falling [9]) (Figure 3). The performance component leads participants through a progressive series of seated postural control tasks (Figure 4). Before testing, participants were provided with written safety instructions. Participants were instructed to engage their wheel locks, and power wheelchair and scooter users were instructed to turn off their devices. All participants were asked to have a handrail or wall nearby in case they lost their balance. To complete the testing, the device guided participants through the completion of three 30-second seated balance tasks in a standardized order that increasingly challenged the participant's base of support: an eyes-open balance task, an eyes-closed balance task, and a functional stability boundary task (Figure

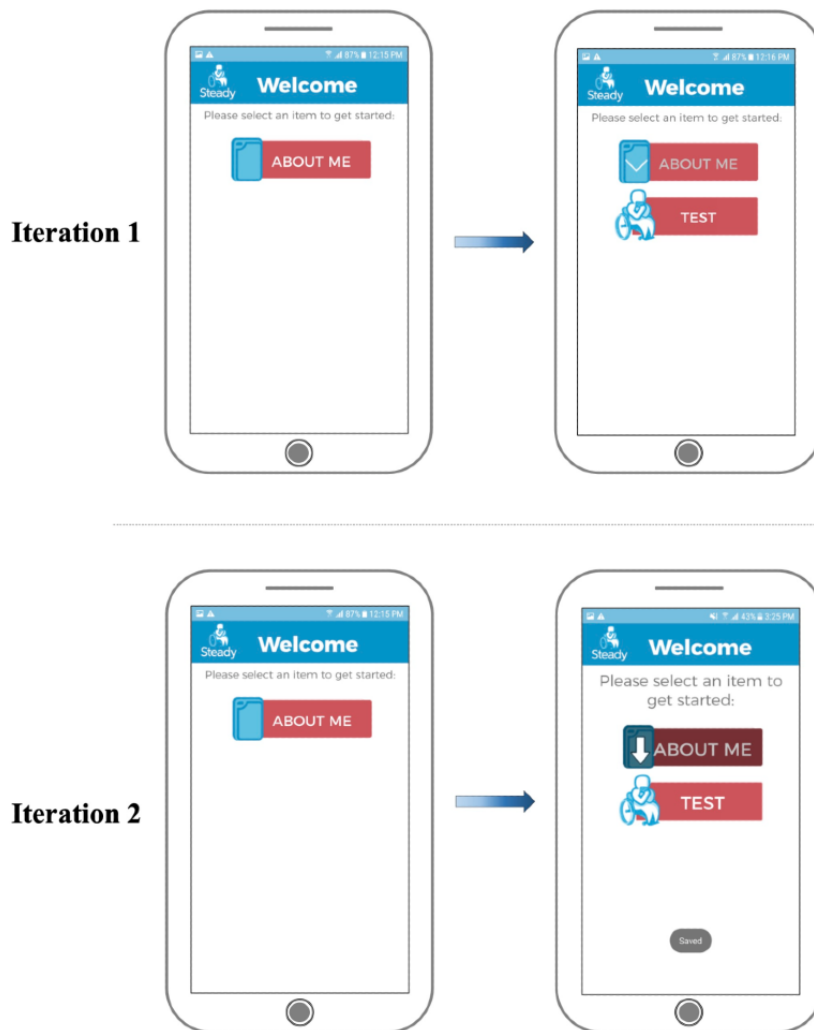
4). These tests were chosen because they can provide insight into postural control [26] and have been linked to fall risk [27]. Written instructions on how to properly complete the balance tasks were provided before the start of each task. After the participant self-selected the "Let's Start" option, an audio and vibratory countdown began from 5, leading to the word "start," which cued the start of the test. The completion of the test was auditorily cued with the word "stop." Participants were asked to hold the smartphone against the middle of their chest with their dominant hand for the duration of each test. Upon completion of each balance task, users reported if they were able to complete the task by selecting one of the following: "I completed the test," "I was unable to complete the test," and "I did NOT attempt to complete the test." If participants were dissatisfied with their attempt at the task, they could select "I'd like to retry" and make another attempt.

A future goal of this work is to utilize the participants' demographic and movement data to generate a personalized fall risk score. To better understand users' preferences for receiving their fall risk score, they were asked to rate different result screen options and provide insight on what made some illustrations better than others (Figure 5).

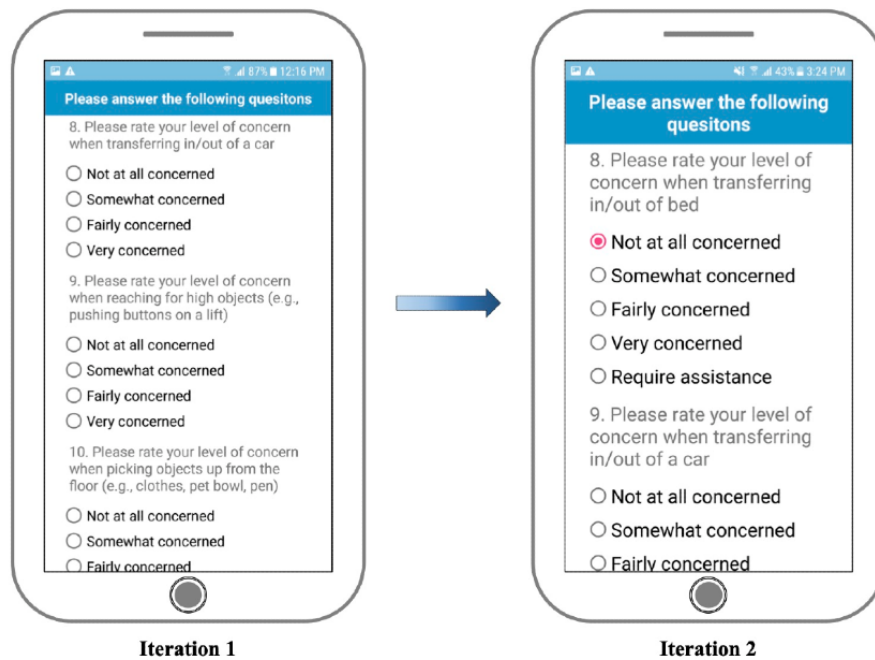
**Figure 1.** The text size was increased and the content was modified from iteration 1 to iteration 2 to allow for greater ease of use.



**Figure 2.** Onscreen instructions were enhanced from iteration 1 to iteration 2.



**Figure 3.** Changes made from iteration 1 to iteration 2 within the “About Me” section included larger text, larger radio buttons, and more choice response options.



**Figure 4.** Modifications were made to the onscreen balance task instructions from iteration 1 to iteration 2.

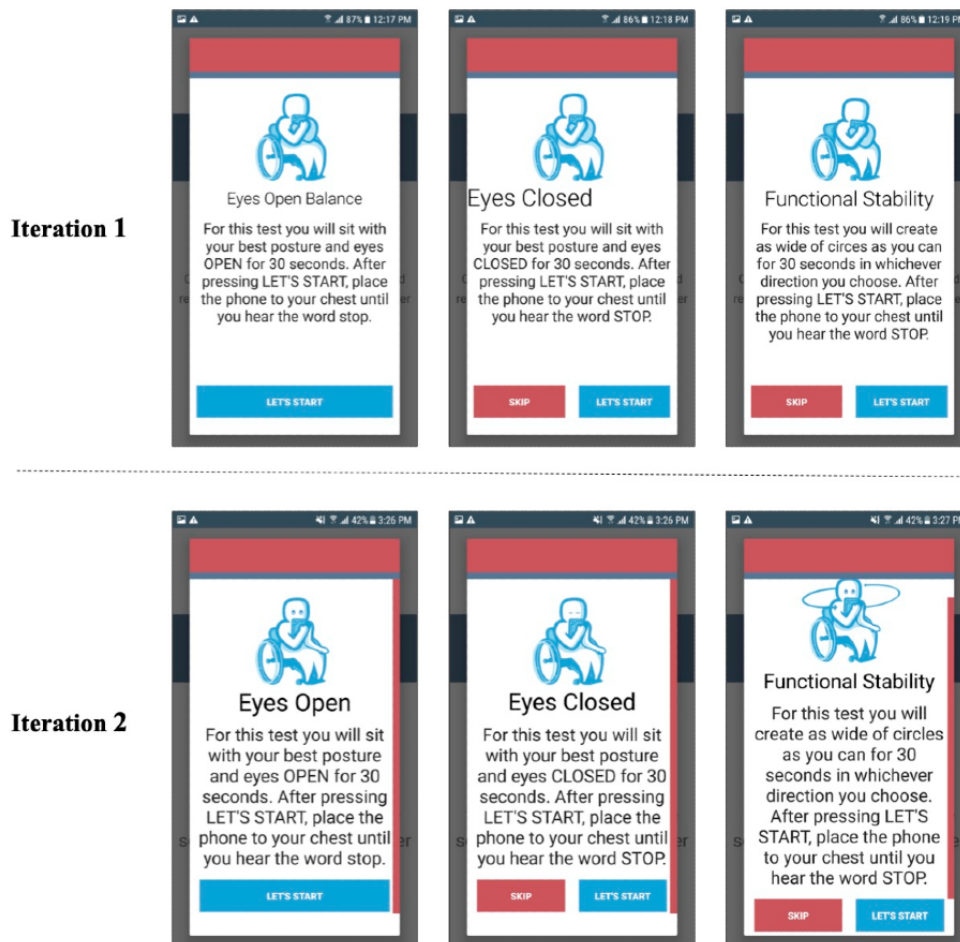


Figure 5. Different options for the result screen (A-E), with modifications made from iteration 1 to iteration 2.



**Ethics Approval**

The Institutional Review Board of the University of Illinois at Urbana-Champaign approved all procedures (20192), and all participants provided informed consent before engaging in research activities. All research procedures were performed in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975.

**Participant Characteristics**

To be eligible, individuals were required to be ≥18 years old, utilize a wheeled mobility device for their main form of mobility, be able to sit unsupported for 30 seconds, have manual dexterity sufficient to use a smartphone, have hearing and vision that were normal or corrected to normal, and be able to read and speak English. In light of the COVID-19 pandemic, having access to video conferencing software (eg, Zoom, Facetime, or Skype), was an inclusion criterion for the second round of testing.

This study included 2 rounds of 5 different older adult wheelchair users (age 58.5 years, SD 12.6 years; 5 male, 5 female) who were recruited from the community through existing participant pools, sharing of research flyers, and word of mouth (Table 1). The first round of testing was completed in person between November 2020 and February 2021, while the second round was performed remotely between April and May 2021. During sessions, participants completed using the app, identified barriers to usability, and gave their rationale for their preferred results options. Feedback from the first round of 5 participants was used to modify the app. Following modification, the second round of volunteers participated in usability testing. This iterative design process is ideal for identifying use challenges, and having a sample size of 5

individuals per round of testing has been shown to be sufficient for identifying usability problems [28,29]. On average, the first round of testing in iterative design identifies 85% of usability problems, and the second round identifies an additional 13% [30]. This approach has been successful in the development of various health apps [31-33], including 2 recent fall risk apps for older adults [18] and patients with multiple sclerosis [34].

**Experimental Session**

After providing informed consent, each participant was given a smartphone (Samsung Galaxy S6, Samsung) that had the Steady-Wheels app installed. The participants were read an instructional prompt (Multimedia Appendix 1) asking them to speak their thought processes aloud while they independently used the app [35]. After the questions were answered, the researchers began visually and auditorily recording the participants’ interactions with the app and wrote field notes. After they completed using the app, the participants completed a semistructured interview in which they were asked to expand upon their likes and dislikes about the app’s layout and features (eg, graphics and wording) and to provide any suggestions for future iterations of the app. During this time, the participants also ranked the fall risk score results options from most to least favorite (Figure 5).

For the most part, these procedures remained constant for the second round of usability testing. The only difference was that the research supplies were delivered to the participants’ residences and the experimental session was completed over video conferencing software.

Along with feedback from the participant interviews, a smartphone usage questionnaire and the System Usability Scale (SUS) [36] were used to understand the participants’ experiences using smartphone and health apps and to quantify the usability

of Steady-Wheels, respectively. While the questionnaire had a total of 6 choice and written response questions, the SUS consists of 10 questions with 5 response options [36,37], ranging from “strongly agree” (5 points) to “strongly disagree” (1 point). After calculation, results from the SUS range from zero (lowest usability) to 100 (highest usability); technology in general has an average score of 60 [37].

### Qualitative Analysis

A thematic analysis approach was used to conduct the qualitative analysis [38]. Video recordings from the think-aloud activity and interviews were transcribed verbatim. The text was then independently reviewed and assigned codes (eg, instructions, testing duration, and graphics) based on its content using the

software MAXQDA (version 12.3.3; Verbi GMBH). Once codes were reviewed and discussed by 2 authors (MF and KH), they were grouped into themes based on the commonality of the data. The same 2 authors (MF and KH) then deliberated on the main themes to ensure they reflected participant insights as accurately as possible. Both authors had prior experience conducting qualitative analyses.

### Results

Participant demographic information is provided in [Table 1](#). [Table 2](#) presents participant responses as the mean response score (with SD) to each question of the SUS for the first and second iterations.

**Table 2.** Participant responses to each System Usability Scale question for the first and second iterations.

System Usability Scale question	Prompt	First iteration, mean score (SD)	Second iteration, mean score (SD)
1	I think that I would like to use this app frequently.	2.6 (1.7)	3.0 (1.6)
2	I found the app unnecessarily complex.	1.0 (0)	1.0 (0)
3	I thought the app was easy to use.	4.8 (0.4)	4.8 (0.5)
4	I think that I would need the support of a technical person to be able to use this app.	1.3 (1.3)	1.0 (0)
5	I found the various functions in this app were well-integrated.	4.2 (0.8)	4.5 (0.6)
6	I thought there was too much inconsistency in this app.	1.4 (0.5)	1.0 (0)
7	I would imagine that most people would learn to use this app very quickly.	4.8 (0.4)	4.8 (0.5)
8	I found the app very cumbersome to use.	1.2 (0.4)	1.0 (0)
9	I felt very confident using the app.	4.4 (0.5)	5.0 (0)
10	I needed to learn a lot of things before I could get going with this app.	2.0 (1.4)	1.3 (0.5)

### Iteration 1

The first round of usability testing yielded 2 themes: ease of use and flexibility of design. Representative participant quotes concerning these themes are reported throughout the following sections. The quotes are accompanied by participant characteristics (eg, sex, and age). System usability scores ranged from 72.5 to 97.5 and averaged 84.5 (SD 11.4), indicating “excellent” usability [39].

#### Ease of Use

Some participants found the app easy to use, saying it was “...very, very straightforward, very easy. I don’t happen to have much to say because it’s pretty straightforward” (male, 47 years old). Others had difficulty determining the sequence in which to complete the separate modules, stating, “There’s only one item here. It says about me. Is that what I’m supposed to touch?” (male, 72 years old); this module can be seen in [Figure 2](#). Following the completion of the “About Me” section, another participant said, “Now do I do the test?” (female, 43 years old). Although most participants were able to navigate the app, their thought processes indicated unnecessary cognitive load regarding the app layout: “Okay. Now we’re ready to do the test

portion I assume since I filled out the about me, so I’ll go ahead and do that” (male, 47 years old). Such insights may help to explain the large variance in participant responses to SUS question 10, which asks “I needed to learn a lot of things before I could get going with this app” ([Table 2](#)). In response to this feedback, the welcome screen was edited to provide a more thorough description of what the app was going to ask of the participant and the order in which it would be completed ([Figure 1](#)). The transition from the “About Me” section to the “Test” section was made more evident by shading the completed “About Me” option and providing a larger arrow pointing to the “Test” option ([Figure 2](#)).

Further participant feedback indicated that the app could be improved by having larger text and multiple-choice buttons, particularly within the “About Me” section. One participant said, “The layout? I guess I would say that some of it is a little bit small in terms of text and radio buttons. Since you’re really focusing on your design you could blow it up a little...there’s plenty of real estate to play with, so you might as well. Especially given the demographics of the people that will be using it—easier to make it more accessible” (male, 47 years old). [Figure 1](#), [Figure 3](#), and [Figure 4](#) illustrate the changes

made to text and radio button size. This increased font size led to the introduction of a vertical slide bar on slides that no longer fit on a single screen (Figure 4). Seated balance task titles were bolded and centered to draw attention (Figure 4).

The purpose of the app's graphics was to help users further understand the instructional text. Based on participant feedback, it became apparent that the graphics used within the first iteration could be further refined. One participant stated, "It'd be easier if it [the graphics] demonstrated exactly what it was saying" (male, 47 years old). To better depict the nature of the tasks, open and closed eyes were added, an arm was moved to the side of the icon's body to illustrate that only one hand was needed to hold the phone to the chest, and circular arrows were positioned around the icon completing the functional stability boundary task to represent the movement pattern of the task (Figure 4).

### **Flexibility of Design**

Steady-Wheels aims to be applicable to all wheeled device users, but many participants showed difficulty answering the demographic questions accurately, due to the limited choice response options. Participants said, "Level of concern when reaching for higher objects? Well, I would normally ask for help" (male, 72 years old) and "Please rate your level of concern when pushing a wheelchair on uneven surfaces. Well, I don't push my wheelchair anymore" (male, 72 years old). The limited choice response options may have led participants to feel as if the app was not tailored to them, leading to the large variance in participant responses to SUS question 1, which asks, "I think that I would like to use this app frequently" (Table 2). To be more inclusive and comprehensive, more choice response options, such as "require assistance" and "I use a powered device" (Figure 3) were added, in addition to another demographic question asking, "What mobility device do you most commonly use?" with response options of "power wheelchair" or "manual wheelchair."

App features that support individual preferences and allow for easy recovery from errors are known to increase the usability of a system. One feature that participants enjoyed was being able to swipe right to left to progress through the slides and left to right to retrieve prior slides. One participant said "Swiping works. That's useful. In addition to the buttons [eg, "next," "back," and "skip"], swiping left or right seems to work fine" (male, 47 years old). In addition to this, participants had the flexibility to change multiple-choice responses, retrieve prior slides, and reassess balance tasks if they were not pleased with their performance. Another participant said, "Whoops, can I go back? I missed something. It asked me a question" (male, 72 years old). For this participant, having the ability to retrieve prior slides and add or adjust their responses to questions was necessary for the accurate completion of the app. This flexibility of use also helped to counterbalance the difficulties associated with small radio buttons, which we have already discussed.

Primary modifications to the app were to increase the size of text and radio buttons for multiple-choice responses (Figure 1, Figure 3, and Figure 4), improve the on-screen directions (Figure 2 and Figure 4), and add response options to the demographic questions (Figure 3).

### **Preferred Results Screens**

During the first iteration, participants strongly favored the result screen that showed a horizontal scale, numbered from 0 to 100 (Figure 5 C). Positive attributes of this option were the color scheme, the large fall risk number, the brief description of fall risk level, and the horizontal layout. Some criticisms included the lack of upper and lower bounds on the scale, (eg, "72 out of what?" [male, 72 years old]) and lack of clear low, medium, and high cut-off locations on the sliding scale. The participants also felt that a lower fall risk should be represented by a lower number. This feedback informed the development of new results screen options for the second iteration of testing.

### **Iteration 2**

The second round of usability testing yielded 2 themes: app layout and clarity of instruction. SUS scores ranged from 87.5 to 97.5 and averaged 91.9 (SD 4.3), indicating "best imaginable" usability [39].

### **App Layout**

In general, participants were very pleased with the layout of the app during the second iteration of testing. One stated, "I think overall, it's good. I think it's clear" (female, 52 years old) and "I think it was straightforward...it was rather clear, concise, and pretty compact" (male, 42 years old). These improvements may help explain the minimal variance in participant responses to SUS question 10: "I needed to learn a lot of things before I could get going on this app" (Table 2). Although there were no clear modifications that needed to be made to the layout, one participant provided some insight into the app's instructions by stating, "It was all very user friendly, self-explanatory, if you take the time to read it." (male, 53 years old). This statement suggests that the app had high-quality instructions, but perhaps too many of them. Further synthesis of the instructions or the inclusion of visual aids may help alleviate this in future iterations.

### **Clarity of Instructions**

The only instructions that received criticism were the ones for the functional stability boundary test. Despite changes to the visual representation (Figure 4), most participants struggled to understand how to complete the test. One participant said, "Well, I don't know what to do with this one. It says, 'create as wide of circles.' I don't know if that's with my wheelchair, in which case, I'd have to turn it back on. And if I do, I can't hold the phone to my chest." (female, 71 years old). For clarity, the instructional text will be altered to read "For this test, you will create as wide of circles with your trunk as you can..."

### **Preferred Results Screens**

During the second iteration, participants strongly favored the result screen option that showed a dial (Figure 5 E), stating that it was a "real obvious one," and complimenting its representation of low, medium, and high risk. Many related it to their preexisting understanding of a speedometer.

During both iterations, participants enjoyed the simplicity of receiving a single score, stating, "I think having a clear and concise one or two number metric is great. That's perfect" (male, 42 years old). However, most had lingering questions, such as



“How do I use this number?” (male, 72 years old) and “What does it tell me?” after the app was complete. “Maybe give some more information about how the score is actually generated, and maybe give some feedback...maybe having a pop-up recommendation screen at the end for some suggestions with exercises or something like that, might be utilitarian” (male, 42 years old). Further changes to the app are needed to investigate how much information is appropriate and informative for users.

## Discussion

### Principal Results

Understanding the usability of a smartphone app provides insight into the quality and overall satisfaction of the user’s experience. An improved experience could lead to greater use of health apps and increased adherence to suggested interventions [40]. Consequently, the purpose of the current study was to determine the usability of a fall risk mobile health app, Steady-Wheels, and identify key insights into technology development for aging adults who use wheeled devices. Initial design considerations were based on age-related changes and physical limitations associated with disability, including motor, sensation, and cognitive impairments. A mixed-method, iterative design and testing process yielded high SUS scores; the app was rated as having “excellent” and “best imaginable” levels of usability. The main themes for each iteration were informed by participant feedback, with the first round of testing yielding 2 main themes (ease of use and flexibility of design) and the second round of testing yielding 2 different main themes (app layout and clarity of instruction). These themes helped identify insights into app development that could promote usability for aging adults who use wheeled devices.

Overall, participants found that the app was straightforward, easy to use, supportive of individual preferences, and allowed for easy recovery from errors. They appreciated the simple, objective fall risk score. App development and modifications came from participant feedback and insights from previously developed apps [18,34,40-42] and an understanding of usability heuristics for interface design, such as the visibility of the system, use of recognition rather than recall, aesthetics, minimalist design, error prevention, and a match between the system and the real world [43].

The visibility of the system and the timeliness and adequacy of feedback and information to users informed the modifications made to the welcome screen and the overall “step-by-step” style of the app. This approach allowed for the recognition of a recently described task rather than recall of prior instruction. While this promoted the ease of use of Steady-Wheels, the primary complaints about health apps made by users are often related to their aesthetics, especially poor or difficult to interpret color coding, graphics, and fonts [42]. Providing a simple color scheme and font with the inclusion of only essential graphics aided this and created minimal distractions within the app. This approach placed a reduced cognitive load on users and helped to reduce the occurrence of errors.

The primary goal of the results scores at the end of testing was to intuitively convey fall risk results to diverse users. Individuals

learn and retain content better from visual information (eg, cartoons and graphics) [44] and can interpret its meaning much more easily if the design is recognizable [41] or matches real-world experiences and expectations [43]. Providing users with a results option that mimicked their preexisting knowledge of speedometers followed these concepts, was well received, and promoted curiosity in the users about what could be done to lower their fall risk. Collectively, these design features led to the development of an app with high perceived ease of use, which is associated with greater adoption of technology [45].

Along with the personal adoption of technology, it is also important to gain insights into the likelihood of users recommending this technology to other individuals. The SUS has a strong relationship with the Net Promotor Score, which has become a common metric to understand customer loyalty [46]. Consumers will likely promote a product if it achieves a SUS score of 81 or greater [39]. In the current study, both iterations of testing yielded a SUS score above this threshold. This is particularly noteworthy as technology in general has an average SUS score of 60 [37]. Overall, these findings indicate that Steady-Wheels may not only be adopted on a personal level, but will likely also be recommended to others at an equal or greater rate than other forms of technology.

### Lessons Learned

Although this is the first app designed to measure fall risk in aging adults who use wheeled devices, our initial design was informed by the needs of users and learned experiences from prior attempts to develop fall risk screening apps, both for older adults [18] and for people with multiple sclerosis [34]; all of which have received high scores for usability from their respective users. Throughout the iterative testing of Steady-Wheels, we identified key insights that could further inform the development of mobile health apps for older users of wheeled devices. Depending on their physical ability, some individuals are reliant on the use of a single hand for all activities of daily living. Future development of remote assessment should account for this in the test selection and the test’s method of completion. By failing to consider this, researchers may increase the task’s safety risk, complexity, and error rates. It was common to see participants that had difficulties with dexterity, making large buttons to select and swiping options key to the app’s ease of use. Also critical to ease of use was the guided (step-by-step) navigation of the app with clear and brief instructions accompanied by representative illustrations along the way. These features helped to reduce the risk of errors, but if mistakes were made during testing, simple error recovery (eg, allowing for retests or access to previous slides to adjust choice responses) should be made possible. Lastly, intuitive fall risk reporting was achieved through the presentation of a single number located on a color-coordinated continuum for low, medium, and high fall risk. While participants found the simple reporting of their fall risk score to be useful, they were eager to learn ways to improve their fall risk. Providing follow-up preventative information may increase the app’s usefulness and encourage further engagement with the app and shared content [40]. Personalized messaging is an easy and effective strategy for altering patient behavior [47] and could be a feasible way to share such information.

Due to the COVID-19 pandemic, the second iteration of this study was completed remotely. This successful experience highlighted the potential feasibility of the home use of the app. Participants received a smartphone that they may not have been previously exposed to, but were able to turn it on, locate the previously installed app, and follow the instructions to completion. The validity and reliability of this novel measurement tool will need to be measured and compared to common clinical tests [48,49].

### Limitations

The current investigation has three primary limitations: (1) baseline interviews with target users were not conducted to inform the app's initial iteration, (2) all participants in the second round of usability testing were required to have access to videoconferencing software (eg, Zoom, Skype, or Facetime), and (3) 9 of the 10 participants had received some form of higher education.

While the current app considered our target users' characteristics and abilities and the lessons learned during the development of previous fall risk apps, baseline interviews were not performed. Taking a more traditional user-centered approach would likely have highlighted additional thoughts, wants, and needs concerning technology and the app's design. Identifying these key insights early on would have been a way to better serve the target users and help ensure that the researchers' time and resources were being used most efficiently.

Although the second round of usability testing helped to provide insights into the app's feasibility in a home setting, exclusively enrolling individuals that already used videoconferencing software may have created a biased, "technology-friendly"

sample. Unfortunately, this was the only possible method of testing during the COVID-19 pandemic. Moving forward, researchers should aim to prioritize in-person data collection sessions when possible.

Despite efforts to recruit through a variety of methods and locations, most participants had received some form of higher education. Higher education may provide individuals with more experience engaging with technology and a better understanding of it, and our participants may have been more likely to understand the fall risk scores as presented. This, too, may have contributed to bias. Future researchers should consider accounting for this effect by enrolling roughly equal proportions of individuals with different educational backgrounds.

### Conclusions

Previous literature has demonstrated that falls are common for individuals who use wheeled devices and are detrimental. The development of an objective, remote fall risk assessment tool could allow for accessible fall risk screening. Smartphone technology is a promising way to provide users with this information. Overall, aging adults who use wheeled devices found the mobile health app easy to use with a high level of usability due to characteristics such as guided navigation of the app, large text and radio buttons, clear and brief instructions that were accompanied by representative illustrations, and simple error recovery. Intuitive fall risk reporting was achieved through the presentation of a single number on a continuum of colors indicating low, medium, and high risk. Future apps developed for fall risk reporting for this population should consider leveraging the insights identified here to maximize usability.

---

### Acknowledgments

This work was supported by the National Institute on Disability, Independent Living, and Rehabilitation Research (grant 90REGE0006-01-00). MF was funded by the MS Run the US Scholarship.

---

### Authors' Contributions

MF contributed to participant recruitment, data collection, data and statistical analyses, and writing the manuscript. JF contributed to app development and writing the manuscript. KH contributed to data and statistical analyses and writing the manuscript. LR contributed to funding acquisition, experimental design, and writing the manuscript. JS contributed to funding acquisition, experimental design, and writing the manuscript.

---

### Conflicts of Interest

JS declares ownership in Sosnoff Technologies, LLC. Other authors declare no conflicts of interest.

---

Multimedia Appendix 1

Instructional Prompt.

[[DOCX File, 13 KB - formative\\_v6i9e32453\\_app1.docx](#)]

---

### References

1. U.S. Wheelchair User Statistics. Pants Up Easy. URL: <https://www.pantsupeasy.com/u-s-wheelchair-user-statistics/> [accessed 2020-09-17]
2. Karmarkar AM, Dicianno BE, Cooper R, Collins DM, Matthews JT, Koontz A, et al. Demographic profile of older adults using wheeled mobility devices. *J Aging Res* 2011;2011:560358 [FREE Full text] [doi: [10.4061/2011/560358](https://doi.org/10.4061/2011/560358)] [Medline: [21748007](https://pubmed.ncbi.nlm.nih.gov/21748007/)]

3. Colby SL, Ortman JM. Projections of the Size and Composition of the U.S. Population: 2014 to 2060. United States Census Bureau. 2015 Mar. URL: <https://www.census.gov/content/dam/Census/library/publications/2015/demo/p25-1143.pdf> [accessed 2020-09-17]
4. Edwards K, McCluskey A. A survey of adult power wheelchair and scooter users. *Disabil Rehabil Assist Technol* 2010;5(6):411-419. [doi: [10.3109/17483101003793412](https://doi.org/10.3109/17483101003793412)] [Medline: [20450455](https://pubmed.ncbi.nlm.nih.gov/20450455/)]
5. Berg K, Hines M, Allen S. Wheelchair users at home: few home modifications and many injurious falls. *Am J Public Health* 2002 Jan;92(1):48. [doi: [10.2105/ajph.92.1.48](https://doi.org/10.2105/ajph.92.1.48)] [Medline: [11772759](https://pubmed.ncbi.nlm.nih.gov/11772759/)]
6. Kirby RL, Ackroyd-Stolarz SA, Brown MG, Kirkland SA, MacLeod DA. Wheelchair-related accidents caused by tips and falls among noninstitutionalized users of manually propelled wheelchairs in Nova Scotia. *Am J Phys Med Rehabil* 1994;73(5):319-330. [doi: [10.1097/00002060-199409000-00004](https://doi.org/10.1097/00002060-199409000-00004)] [Medline: [7917161](https://pubmed.ncbi.nlm.nih.gov/7917161/)]
7. Rice L, Kalron A, Berkowitz S, Backus D, Sosnoff J. Fall prevalence in people with multiple sclerosis who use wheelchairs and scooters. *Medicine (Baltimore)* 2017 Sep;96(35):e7860 [FREE Full text] [doi: [10.1097/MD.00000000000007860](https://doi.org/10.1097/MD.00000000000007860)] [Medline: [28858096](https://pubmed.ncbi.nlm.nih.gov/28858096/)]
8. Rice LA, Ousley C, Sosnoff JJ. A systematic review of risk factors associated with accidental falls, outcome measures and interventions to manage fall risk in non-ambulatory adults. *Disabil Rehabil* 2015;37(19):1697-1705. [doi: [10.3109/09638288.2014.976718](https://doi.org/10.3109/09638288.2014.976718)] [Medline: [25354146](https://pubmed.ncbi.nlm.nih.gov/25354146/)]
9. Boswell-Ruys CL, Harvey LA, Delbaere K, Lord SR. A Falls Concern Scale for people with spinal cord injury (SCI-FCS). *Spinal Cord* 2010 Sep;48(9):704-709. [doi: [10.1038/sc.2010.1](https://doi.org/10.1038/sc.2010.1)] [Medline: [20125107](https://pubmed.ncbi.nlm.nih.gov/20125107/)]
10. Tinetti ME, Williams CS. The effect of falls and fall injuries on functioning in community-dwelling older persons. *J Gerontol A Biol Sci Med Sci* 1998 Mar;53(2):M112-M119. [doi: [10.1093/gerona/53a.2.m112](https://doi.org/10.1093/gerona/53a.2.m112)] [Medline: [9520917](https://pubmed.ncbi.nlm.nih.gov/9520917/)]
11. Sarmiento K, Lee R. STEADI: CDC's approach to make older adult fall prevention part of every primary care practice. *J Safety Res* 2017 Dec;63:105-109 [FREE Full text] [doi: [10.1016/j.jsr.2017.08.003](https://doi.org/10.1016/j.jsr.2017.08.003)] [Medline: [29203005](https://pubmed.ncbi.nlm.nih.gov/29203005/)]
12. Roeing KL, Hsieh KL, Sosnoff JJ. A systematic review of balance and fall risk assessments with mobile phone technology. *Arch Gerontol Geriatr* 2017 Nov;73:222-226. [doi: [10.1016/j.archger.2017.08.002](https://doi.org/10.1016/j.archger.2017.08.002)] [Medline: [28843965](https://pubmed.ncbi.nlm.nih.gov/28843965/)]
13. Reyes A, Qin P, Brown CA. A standardized review of smartphone applications to promote balance for older adults. *Disabil Rehabil* 2018 Mar;40(6):690-696. [doi: [10.1080/09638288.2016.1250124](https://doi.org/10.1080/09638288.2016.1250124)] [Medline: [27868438](https://pubmed.ncbi.nlm.nih.gov/27868438/)]
14. Mellone S, Tacconi C, Schwickert L, Klenk J, Becker C, Chiari L. Smartphone-based solutions for fall detection and prevention: the FARSEEING approach. *Z Gerontol Geriatr* 2012 Dec;45(8):722-727. [doi: [10.1007/s00391-012-0404-5](https://doi.org/10.1007/s00391-012-0404-5)] [Medline: [23184298](https://pubmed.ncbi.nlm.nih.gov/23184298/)]
15. Rasche P, Mertens A, Bröhl C, Theis S, Seinsch T, Wille M, et al. The "Aachen fall prevention App" - a Smartphone application app for the self-assessment of elderly patients at risk for ground level falls. *Patient Saf Surg* 2017;11:14 [FREE Full text] [doi: [10.1186/s13037-017-0130-4](https://doi.org/10.1186/s13037-017-0130-4)] [Medline: [28503199](https://pubmed.ncbi.nlm.nih.gov/28503199/)]
16. Wetsman N. Apple's new health features bring new focus to elder care technology. *The Verge*. URL: <https://www.theverge.com/2021/6/10/22527707/apple-health-data-elderly-falls-walking-privacy> [accessed 2021-07-06]
17. Hsieh KL, Roach KL, Wajda DA, Sosnoff JJ. Smartphone technology can measure postural stability and discriminate fall risk in older adults. *Gait Posture* 2019 Jan;67:160-165. [doi: [10.1016/j.gaitpost.2018.10.005](https://doi.org/10.1016/j.gaitpost.2018.10.005)] [Medline: [30340129](https://pubmed.ncbi.nlm.nih.gov/30340129/)]
18. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A Fall Risk mHealth App for Older Adults: Development and Usability Study. *JMIR Aging* 2018 Nov 20;1(2):e11569 [FREE Full text] [doi: [10.2196/11569](https://doi.org/10.2196/11569)] [Medline: [31518234](https://pubmed.ncbi.nlm.nih.gov/31518234/)]
19. Frechette ML, Abou L, Rice LA, Sosnoff JJ. The Validity, Reliability, and Sensitivity of a Smartphone-Based Seated Postural Control Assessment in Wheelchair Users: A Pilot Study. *Front Sports Act Living* 2020;2:540930 [FREE Full text] [doi: [10.3389/fspor.2020.540930](https://doi.org/10.3389/fspor.2020.540930)] [Medline: [33367273](https://pubmed.ncbi.nlm.nih.gov/33367273/)]
20. 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 05;16(6):e146 [FREE Full text] [doi: [10.2196/jmir.3077](https://doi.org/10.2196/jmir.3077)] [Medline: [24905070](https://pubmed.ncbi.nlm.nih.gov/24905070/)]
21. Feys P, Lamers I, Francis G, Benedict R, Phillips G, LaRocca N, Multiple Sclerosis Outcome Assessments Consortium. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler* 2017 Apr;23(5):711-720 [FREE Full text] [doi: [10.1177/1352458517690824](https://doi.org/10.1177/1352458517690824)] [Medline: [28206826](https://pubmed.ncbi.nlm.nih.gov/28206826/)]
22. Uzochukwu JC, Stegemöller EL. Repetitive Finger Movement and Dexterity Tasks in People With Parkinson's Disease. *Am J Occup Ther* 2019;73(3):7303205090p1-7303205090p8. [doi: [10.5014/ajot.2019.028738](https://doi.org/10.5014/ajot.2019.028738)] [Medline: [31120839](https://pubmed.ncbi.nlm.nih.gov/31120839/)]
23. Johansson GM, Häger CK. A modified standardized nine hole peg test for valid and reliable kinematic assessment of dexterity post-stroke. *J Neuroeng Rehabil* 2019 Jan 14;16(1):8 [FREE Full text] [doi: [10.1186/s12984-019-0479-y](https://doi.org/10.1186/s12984-019-0479-y)] [Medline: [30642350](https://pubmed.ncbi.nlm.nih.gov/30642350/)]
24. Erol AM, Ceceli E, Uysal Ramadan S, Borman P. Effect of rheumatoid arthritis on strength, dexterity, coordination and functional status of the hand: the relationship with magnetic resonance imaging findings. *Acta Reumatol Port* 2016;41(4):328-337 [FREE Full text] [Medline: [27926913](https://pubmed.ncbi.nlm.nih.gov/27926913/)]
25. Bernard M, Liao C, Mills M. The effects of font type and size on the legibility and reading time of online text by older adults. In: CHI '01 Extended Abstracts on Human Factors in Computing Systems. 2001 Presented at: CHI01: Human Factors in Computing Systems; Mar 31-Apr 5, 2001; Seattle, WA p. 175-176. [doi: [10.1145/634067.634173](https://doi.org/10.1145/634067.634173)]

26. Shin S, Sosnoff JJ. Spinal cord injury and time to instability in seated posture. *Arch Phys Med Rehabil* 2013 Aug;94(8):1615-1620. [doi: [10.1016/j.apmr.2013.02.008](https://doi.org/10.1016/j.apmr.2013.02.008)] [Medline: [23422406](https://pubmed.ncbi.nlm.nih.gov/23422406/)]
27. Sung J, Ousley CM, Shen S, Isaacs ZJK, Sosnoff JJ, Rice LA. Reliability and validity of the function in sitting test in nonambulatory individuals with multiple sclerosis. *Int J Rehabil Res* 2016 Dec;39(4):308-312. [doi: [10.1097/MRR.000000000000188](https://doi.org/10.1097/MRR.000000000000188)] [Medline: [27442603](https://pubmed.ncbi.nlm.nih.gov/27442603/)]
28. How Many Test Users in a Usability Study? Nielsen Norman Group. URL: <https://www.nngroup.com/articles/how-many-test-users/> [accessed 2020-09-18]
29. Macefield R. How To Specify the Participant Group Size for Usability Studies: A Practitioner's Guide. *J Usability Stud* 2009;5(1):34-45 [FREE Full text]
30. Why You Only Need to Test with 5 Users. Nielsen Norman Group. URL: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> [accessed 2022-04-22]
31. Wu Y, Liu H, Kelleher A, Pearlman J, Cooper RA. Evaluating the usability of a smartphone virtual seating coach application for powered wheelchair users. *Med Eng Phys* 2016 Jun;38(6):569-575. [doi: [10.1016/j.medengphy.2016.03.001](https://doi.org/10.1016/j.medengphy.2016.03.001)] [Medline: [27079179](https://pubmed.ncbi.nlm.nih.gov/27079179/)]
32. Zou P, Stinson J, Parry M, Dennis C, Yang Y, Lu Z. A Smartphone App (mDASHNa-CC) to Support Healthy Diet and Hypertension Control for Chinese Canadian Seniors: Protocol for Design, Usability and Feasibility Testing. *JMIR Res Protoc* 2020 Apr 02;9(4):e15545 [FREE Full text] [doi: [10.2196/15545](https://doi.org/10.2196/15545)] [Medline: [32238343](https://pubmed.ncbi.nlm.nih.gov/32238343/)]
33. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. *J Med Internet Res* 2013 Mar 08;15(3):e51 [FREE Full text] [doi: [10.2196/jmir.2350](https://doi.org/10.2196/jmir.2350)] [Medline: [23475457](https://pubmed.ncbi.nlm.nih.gov/23475457/)]
34. Hsieh K, Fanning J, Frechette M, Sosnoff J. Usability of a Fall Risk mHealth App for People With Multiple Sclerosis: Mixed Methods Study. *JMIR Hum Factors* 2021 Mar 22;8(1):e25604 [FREE Full text] [doi: [10.2196/25604](https://doi.org/10.2196/25604)] [Medline: [33749609](https://pubmed.ncbi.nlm.nih.gov/33749609/)]
35. Fonteyn ME, Kuipers B, Grobe SJ. A Description of Think Aloud Method and Protocol Analysis. *Qual Health Res* 2016 Jul 01;3(4):430-441. [doi: [10.1177/104973239300300403](https://doi.org/10.1177/104973239300300403)]
36. Brooke J. SUS - A quick and dirty usability scale. Redhatch Consulting Ltd. URL: <https://hell.meiert.org/core/pdf/sus.pdf> [accessed 2020-09-21]
37. Measuring Usability with the System Usability Scale (SUS). MeasuringU. URL: <https://measuringu.com/sus/> [accessed 2020-09-21]
38. Creswell JW. *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publications; 1998:xv, 403.
39. 5 Ways to Interpret a SUS Score. MeasuringU. URL: <https://measuringu.com/interpret-sus-score/> [accessed 2020-11-30]
40. Georgsson M, Staggers N. An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. *J Biomed Inform* 2016 Feb;59:115-129 [FREE Full text] [doi: [10.1016/j.jbi.2015.11.008](https://doi.org/10.1016/j.jbi.2015.11.008)] [Medline: [26639894](https://pubmed.ncbi.nlm.nih.gov/26639894/)]
41. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Design, development, and formative evaluation of a smartphone application for recording and monitoring physical activity levels: the 10,000 Steps "iStepLog". *Health Educ Behav* 2013 Apr;40(2):140-151. [doi: [10.1177/1090198112449460](https://doi.org/10.1177/1090198112449460)] [Medline: [22984196](https://pubmed.ncbi.nlm.nih.gov/22984196/)]
42. 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](https://doi.org/10.1016/j.jbi.2016.02.002)] [Medline: [26903153](https://pubmed.ncbi.nlm.nih.gov/26903153/)]
43. 10 Heuristics for User Interface Design Internet. Nielsen Norman Group. URL: <https://www.nngroup.com/articles/ten-usability-heuristics/> [accessed 2021-10-23]
44. Burmark L. *Visual Literacy: Learn To See, See To Learn*. Arlington, VA: Association for Supervision & Curriculum Development; 2002.
45. 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](https://doi.org/10.2307/249008)]
46. Predicting Net Promoter Scores from System Usability Scale Scores. MeasuringU. URL: <https://measuringu.com/nps-sus/> [accessed 2020-11-30]
47. Fried TR, Redding CA, Robbins ML, Paiva AL, O'Leary JR, Iannone L. Development of Personalized Health Messages to Promote Engagement in Advance Care Planning. *J Am Geriatr Soc* 2016 Feb;64(2):359-364 [FREE Full text] [doi: [10.1111/jgs.13934](https://doi.org/10.1111/jgs.13934)] [Medline: [26804791](https://pubmed.ncbi.nlm.nih.gov/26804791/)]
48. Gorman SL, Radtka S, Melnick ME, Abrams GM, Byl NN. Development and validation of the Function In Sitting Test in adults with acute stroke. *J Neurol Phys Ther* 2010 Sep;34(3):150-160. [doi: [10.1097/NPT.0b013e3181f0065f](https://doi.org/10.1097/NPT.0b013e3181f0065f)] [Medline: [20716989](https://pubmed.ncbi.nlm.nih.gov/20716989/)]
49. Abou L, Sung J, Sosnoff JJ, Rice LA. Reliability and validity of the function in sitting test among non-ambulatory individuals with spinal cord injury. *J Spinal Cord Med* 2020 Nov;43(6):846-853 [FREE Full text] [doi: [10.1080/10790268.2019.1605749](https://doi.org/10.1080/10790268.2019.1605749)] [Medline: [30998421](https://pubmed.ncbi.nlm.nih.gov/30998421/)]

---

**Abbreviations**

**SUS:** System Usability Scale

---

---

*Edited by T Leung; submitted 29.07.21; peer-reviewed by P Rasche, A Joseph; comments to author 04.01.22; revised version received 22.04.22; accepted 27.06.22; published 16.09.22.*

*Please cite as:*

*Frechette M, Fanning J, Hsieh K, Rice L, Sosnoff J*

*The Usability of a Smartphone-Based Fall Risk Assessment App for Adult Wheelchair Users: Observational Study*

*JMIR Form Res 2022;6(9):e32453*

*URL: <https://formative.jmir.org/2022/9/e32453>*

*doi: [10.2196/32453](https://doi.org/10.2196/32453)*

*PMID: [36112405](https://pubmed.ncbi.nlm.nih.gov/36112405/)*

©Mikaela Frechette, Jason Fanning, Katherine Hsieh, Laura Rice, Jacob Sosnoff. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 16.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Acceptability and Impact of an Educational App (iCare) for Informal Carers Looking After People at Risk of Pressure Ulceration: Mixed Methods Pilot Study

Eamonn McKeown<sup>1</sup>, PhD; Caroline McGraw<sup>1</sup>, PhD; Pru Holder<sup>1</sup>, MSc; Jenny Shand<sup>2</sup>, PhD; Shashivadan P Hirani<sup>1</sup>, PhD

<sup>1</sup>School of Health & Psychological Sciences, City, University of London, London, United Kingdom

<sup>2</sup>Faculty of Brain Sciences, University College London, London, United Kingdom

**Corresponding Author:**

Eamonn McKeown, PhD

School of Health & Psychological Sciences

City, University of London

Northampton Square

London, EC1V 0HB

United Kingdom

Phone: 44 70405917

Email: [eamonn.mckeown.1@city.ac.uk](mailto:eamonn.mckeown.1@city.ac.uk)

## Abstract

**Background:** Pressure ulcers are areas of skin damage resulting from sustained pressure. Informal carers play a central role in preventing pressure ulcers among older and disabled people living at home. Studies highlight the paucity of pressure ulcer training for informal carers and suggest that pressure ulcer risk is linked to high levels of carer burden.

**Objective:** This pilot study evaluated a smartphone app with a specific focus on pressure ulcer prevention education for informal carers. The app was developed based on the principles of microlearning. The study aimed to explore carer perspectives on the acceptability of the app and determine whether the app increased knowledge and confidence in their caring role.

**Methods:** In this concurrent mixed methods study, participants completed quantitative questionnaires at baseline and at the end of weeks 2 and 6, which examined caregiving self-efficacy, preparedness for caregiving, caregiver strain, pressure ulcer knowledge, and app acceptability and usability. A subsample of participants participated in a “think aloud” interview in week 1 and semistructured interviews at the end of weeks 2 and 6.

**Results:** Of the 32 participants, 23 (72%) participants completed the week 2 and 16 (50%) completed the week 6 questionnaires; 66% (21/32) of carers participated in qualitative “think aloud” interviews, and 18 (56%) also participated in semistructured interviews at week 2, and 13 (41%) at week 6. Pressure ulcer knowledge scores significantly changed ( $F_{1,6.112}=21.624$ ;  $P=.001$ ) from baseline (mean 37.5; SE 2.926) to the second follow-up (mean 59.72, SE 3.985). Regarding the qualitative data, the theme “I’m more careful now and would react to signs of redness” captured participants’ reflections on the new knowledge they had acquired, the changes they had made to their caring routines, their increased vigilance for signs of skin damage, and their intentions toward the app going forward. There were no significant results pertaining to improved preparedness for caregiving or caregiving self-efficacy or related to the Caregiver Strain Index. Participants reported above average usability scores on a scale of 0 to 100 (mean 69.94, SD 18.108). The app functionality and information quality were also rated relatively high on a scale of 0 to 5 (mean 3.84, SD 0.704 and mean 4.13, SD 0.452, respectively). Overall, 2 themes pertaining to acceptability and usability were identified: “When you’re not used to these things, they take time to get the hang of” and “It’s not a fun app but it is informative.” All participants ( $n=32$ , 100%) liked the microlearning approach.

**Conclusions:** The iCare app offers a promising way to improve informal carers’ pressure ulcer knowledge. However, to better support carers, the findings may reflect the need for future iterations of the app to use more interactive elements and the introduction of gamification and customization based on user preferences.

(JMIR Form Res 2022;6(9):e36517) doi:[10.2196/36517](https://doi.org/10.2196/36517)

## KEYWORDS

pressure ulcers; informal carers; smartphone apps; mobile health; mHealth; educational technology; health education; mobile phone

## Introduction

### Background

Pressure ulcers are defined as localized damage to the skin or underlying tissue, typically over a bony prominence, resulting from sustained pressure, which may present as intact skin or open ulcer [1]. They are usually caused by prolonged sitting or lying in one position. Populations at high risk include people who are frail and old [2] and people with spinal cord injuries [3]. According to a cross-sectional study, the prevalence of pressure ulcers is between 0.40 and 0.77 per 1000 community-dwelling adults in England [4].

Pressure ulcers are a substantial source of burden. They cause pain, exudate, and odor [5], and affect a person's ability to participate in rehabilitation [6]. Pressure ulcers are also slow to heal. Most are treated using dressings and topical agents; however, some require surgical repair. Complications include soft-tissue and bone infections. Infection can cause sepsis and even death. Annually, the United Kingdom National Health Service (NHS) treats 202,000 people for pressure ulcers, at a yearly cost of £571.98 million (US \$660.23 million) [7].

The essential elements of pressure ulcer prevention and management are the following: providing appropriate support surfaces (eg, pressure relieving cushions and mattresses), conducting regular skin inspections, supporting patients to keep moving, ensuring that incontinence and moisture are managed, and maintaining adequate nutrition and hydration [8]. Regarding nursing management, depending on ulcer severity, most people with pressure ulcers receive between 1 and 3 nursing visits per week for wound care [9].

Studies exploring the factors influencing the implementation of evidence-based practice in pressure ulcer prevention and management in community settings have identified how health care practitioners regard informal carers as central to both pressure ulcer prevention and wound healing [10]. In total, 4 recent studies have explored carer input in pressure ulcer prevention and management and their perspectives of the factors affecting pressure ulcer care in home settings [11-14]. The findings emphasized high levels of carer burden and highlighted the paucity of carer training and the importance of communication with health care practitioners.

### Supporting Carers

In this study, the term "informal carer" is defined as someone providing unpaid care to an older dependent person where there is an existing social relationship (eg, a spouse or other relative). All subsequent references to carers will be to those working in this informal capacity. In 2015, NHS England pledged to raise the profile of such carers and how to support them [15]. Recent systematic reviews suggest that web-based interventions may result in a range of improved health outcomes for carers, including reductions in depression, stress, anxiety, social isolation, and relationship problems [16,17]. Moreover, these

studies have suggested that robust web-mediated carer education has the potential to enhance management of the caring role with a concomitant reduction in the requirement for health care practitioner input.

With improved accessibility of smartphone devices, the role of smartphone health care apps is expanding. Smartphone apps can support carers by providing access to information, support, and resources at any time where the person has internet connection. App information is also easier to update and alert users to (via a notification in the app) than equivalent paper versions, which will be more expensive and difficult to ensure that users are reached. Previous studies have assessed the use of health care apps among carers of people with cancer [18] and carers of people living with dementia [19]. Although these studies found a positive attitude toward apps among carers [18,19], they also identified barriers to their use including concerns about time constraints and not being familiar with technology [19].

### Contribution of This Study

To the best of our knowledge, no health care app has been evaluated among carers of people at risk of pressure ulceration. The aim of our study was to (1) explore carer perspectives on the acceptability and usability of a pressure ulcer app and (2) determine whether the app increased carers' knowledge and confidence in their caring role.

## Methods

### Design

This was a 6-week, concurrent mixed methods pilot study in which participants were given access to a smartphone education app which had specific focus on carers and the care, management, and prevention of pressure ulcers. The study involved two components: (1) web-based Qualtrics-based questionnaires, completed by carers in weeks 1, 2, and 6, and (2) "think aloud" interviews with carers in week 1 and semistructured interviews with carers in weeks 2 and 6. The Good Reporting of a Mixed Methods Study guidelines [20] were adhered to in the reporting of this study.

### Ethics Approval

Ethics approval for the study was granted by the School of Health Sciences Research Ethics Committee at City, University of London (ETH1819-1600), and relevant governance approvals were received from the local NHS provider organization.

### iCare App

The app was developed by Care City, a Community Interest Company, which aims to work with residents and organizations to improve health and well-being in Northeast London, by bringing together health, social, and third sector partners; technology experts; and researchers. Using the Agylia Learning Management System, the app design and format were shaped by the principles of microlearning, in which short, focused

pieces of content are provided to an audience, when and where they need it [21]. The app's content was organized into 14 units, with each unit consisting of a video presentation, written information, and interactive learning objects (Table 1). Each of the learning units was designed to take approximately 3 minutes

to complete. The content of the app reflected information contained within an educational pack for carers that was developed by the local NHS provider organization [22]. The app was available to download on iPhone and Android devices.

**Table 1.** iCare app—learning unit topics and unit format.

Unit number	Learning unit topic	Learning unit format
1	What are pressure ulcers?	Video (2 minutes 46 seconds)
2	Frequently asked questions about pressure ulcers	Video (40 seconds), written frequently asked questions, and pictures of pressure ulcers
3	Five things you should know about keeping people moving	Bullet-pointed list and interactive components
4	Five things you should know about keeping skin healthy	Bullet-pointed list and components
5	Five things you should know about nutrition	Bullet-pointed list and interactive components
6	Five things you should know about support surfaces	Bullet-pointed list and interactive components
7	How to ensure adequate nutrition?	Video (1 minute 19 seconds) and interactive components
8	How to help people keep moving?	Video (1 minute 40 seconds) and interactive components
9	How to keep skin healthy?	Video (1 minute 27 seconds) and interactive components
10	How to support people at risk effectively?	Video (1 minute 37 seconds) and interactive components
11	Pressure ulcer triggers	Interactive checklist
12	Skin inspection guide	Interactive checklist
13	Sources of help	Color-coded reference chart
14	Identifying who is at risk of getting a pressure ulcer	Color-coded reference chart

## Sampling and Recruitment

The study was conducted in London, England. Individuals meeting the following inclusion criteria were eligible to participate: (1) aged  $\geq 18$  years, (2) identifiable as an informal carer for a person with or at risk of pressure ulcer, (3) able to participate in the interview in English, and (4) have access to an iPhone or iPad or Android device.

For pilot studies, the sample size for quantitative components is suggested to be 30 [23,24], which will allow parameter estimates and loss to follow-up rates for subsequent large studies. The sample size for the “think aloud” interview and semistructured interviews were influenced by the concept of data saturation [25]. Given the topic area was clearly defined, a sample of 15 participants was expected to achieve data saturation. Previous studies using the “think aloud” approach to usability testing for health care apps have used sample sizes of 10 [26] and 24 [27], respectively.

The study was advertised on posters displayed in public areas on NHS sites in East London (including general practitioner surgery and rehabilitation wards). The study was also promoted by Care City staff attending local carer support group events. At these events, staff explained the purpose, methods, and intended uses of the study. They also explained that, depending on carer preference, participation will entail either (1) the completion of 3 web-based questionnaires or (2) the completion of 3 web-based questionnaires and participation in 1 face-to-face “think aloud” interview and 2 additional semistructured interviews. In total, 14 events were attended, at which there

were approximately 150 carers; however, not all attendees met the study inclusion criteria. In total, 29 eligible carers expressed interest in participating and were provided with a participant information sheet, and consent was obtained for their contact details to be shared with both the app registration team and the research team. According to Care City, reported barriers to recruitment included the perceived relevance of pressure ulcers and digital exclusion. Regarding the former, many of those attending the carers events did not consider the person they cared for as being at risk of pressure ulceration, and therefore did not think that the app will be of benefit to them. Regarding digital exclusion, many carers reported that they did not have access to the right technology, whereas others did not feel sufficiently technologically confident to engage with an app.

Following agreement, the app registration team set up individual user accounts and emailed carers their account details and instructions for downloading the app. Only carers who expressed interest in being interviewed as part of the study were referred to the research team, who telephoned them to confirm their ongoing interest and arrange a convenient time and location for the first interview.

## Data Collection

All participants were asked to use the app for a period of 6 weeks. Data were collected between October 2019 and April 2020.



## Web-Based Questionnaires

### Overview

Participation for the whole sample comprised completion of web-based questionnaires at three time points: (1) at the start of week 1, (2) at the end of week 2, and (3) at the end of week 6. The questionnaires were administered via a web-based platform, Qualtrics. The first page of each questionnaire had a consent statement. Participants were directed to complete the questionnaire only after they read the statement and agreed to participate. Participants were prompted to complete the questionnaires via automatic emails sent at the start of week 1 and at the end of weeks 2 and 6. Anyone not completing the questionnaire within 7 days from the specified date received a telephone reminder from the app registration team. Participants received a £5 (US \$5.81) e-voucher after completing each questionnaire to compensate them for their time and effort.

### Week 1 (Baseline)

The baseline questionnaire comprised 3 main sections. In the first section, participants were asked to provide demographic information pertaining to their gender, age, ethnic background, highest level of education, relationship with the care recipient, and previous care-related training and the age and gender of the care recipient and their primary diagnosis. They were also asked whether they had previously used any health app or apps.

The second section measured existing pressure ulcer knowledge using a 20-item questionnaire, which was used to produce 2 parallel forms of 12 items each, at different time points (weeks 1 and 6). Items were generated from the educational pack developed by the local NHS provider organization for carers on how they can support family members at risk of pressure ulceration [22] and the Pressure Ulcer Knowledge Assessment Tool 2.0 questionnaire for registered nurses and nursing assistants [28]. Then, the items were clustered around four categories: (1) support surfaces, (2) nutrition and hydration, (3) keep moving, and (4) skin care and inspection.

The third section measured participants' self-reported outcomes including confidence in dealing with caregiving situations, using the Caregiving Self-Efficacy Scale (CSES) [29]. The choice of responses ranged from 1 (not at all confident) to 5 (extremely confident). This section also measured how prepared participants were for their caregiving role, using the Preparedness for Caregiving Scale (PfCS) [30]. The choice of responses ranged from 0 (not at all prepared) to 4 (very well prepared). Finally, strain related to the caregiving role was measured using the Caregiver Strain Index (CSI) [31].

The week 1 questionnaire took approximately 30 minutes to complete.

### Week 2 (First Follow-up)

This questionnaire comprised only 1 section, in which participants were asked to complete the System Usability Scale (SUS) [32] and the Mobile App Rating Scale (MARS) [33]. The former is a 5-point Likert scale (ranging from 1="strongly disagree" to 5="strongly agree"), giving a global view of subjective assessments of usability. The MARS also uses a

range of Likert-type scale responses. The questionnaire took approximately 10 minutes to complete.

### Week 6 (Second Follow-up)

At the end of week 6, participants were again asked to complete the CSES, PfCS, and CSI and answer follow-up questions on their pressure ulcer knowledge. The questionnaire took approximately 10 minutes to complete.

### "Think Aloud" Interviews

The "think aloud" interviews were conducted with a subgroup of participants at the beginning of week 1. They were conducted face to face in a place with internet access chosen by the participant. The "think aloud" approach [34] was selected on the basis that it will provide a useful reflection on the carers' cognitive processes and attitudes while downloading and using the app for the first time. To gain experience with the think aloud method, the interviewer (PH) conducted 2 pilot interviews—one with someone who had no previous exposure to this approach and one with someone who had extensive experience.

Written consent was obtained from participants before the interview. During the interview, the participant downloaded the iCare app from either the iPhone App Store or the Google Play Store. Then, the participants were encouraged to interact with the content while the interviewer asked them to verbalize their thought processes (eg, to voice any confusion or trouble they were having while navigating the app) and attitudes toward the content. All interviews were audio-recorded with participants' permission. At the end of the interview, the interviewer made an appointment with the participant for their first semistructured interview (refer to the following section).

### Semistructured Interviews

The subgroup that participated in the "think aloud" interviews were invited to participate in one-on-one semistructured interviews at the end of weeks 2 and 6. The topic guide for these interviews asked about participants' use of the iCare app since the previous interview, their perceptions of using the app, changes they had made because of using the app, their plans for continuing to use the app, best thing about the app, how the app may be improved, and what other sources of pressure ulcer information they had accessed since the previous interview. Interviews were conducted by PH via telephone. Written consent was obtained before the interview. Interviews were digitally recorded with participant's permission. At the end of the first semistructured interview, the interviewer made an appointment for the second interview.

### Data Analysis

#### Statistical Analysis

Quantitative data were entered into SPSS and analyzed for (1) description of the sample at baseline; (2) descriptives of the sample's mobile app rating at the first follow-up; (3) relationships between continuous and ordinal variables, using Pearson or Spearman correlations; and (4) changes in outcomes from baseline to second follow-up, using linear mixed models analyses. Descriptive statistics (eg, means and SDs) have been produced and are presented in the Results section. Data were

screened to check whether they met the assumptions of parametric statistics, and appropriate inferential statistics were conducted. The statistical analysis was performed by SH.

### **Qualitative Data Analysis**

Interviews were transcribed verbatim by an independent professional transcription service. Pilot data were not included in the analysis. Data were sifted and interpreted using the framework approach to qualitative data analysis [35], which allows the analytical process to be informed by issues designated in advance and new and emergent themes [36]. In the steps of this approach, transcription is followed by familiarization, coding, analytical framework development, indexing, charting, and interpreting. Deductive coding was guided by the study's aim and used predefined codes derived from the MARS [33], Service User Technology Acceptability Questionnaire [37], and Treatment Acceptability Framework [38]. In total, 2 members of the research team (CM and EM) independently coded a sample of the transcripts. The remaining transcripts were coded by CM, and an analytical framework was developed. After the framework was developed and data were charted into the matrix, the data were interpreted by CM. All interpretations were discussed and interrogated by other members of the research team (EM and SH).

## **Results**

The quantitative and qualitative results are integrated and presented in two parts to meet the aim of the study: (1) acceptability and usability of the iCare app and (2) impact of the iCare app on carers' knowledge and confidence in their caring role.

### **Sample Characteristics**

In total, 32 participants were registered with the iCare app. were carers who had attended one of the aforementioned carers events, or carers who had responded to posters advertising the

study in public areas on NHS sites. Table 2 shows the characteristics of participants.

The mean age of the sample was 57.9 (SD 11.15) years, with 69% (22/32) women and 31% (10/32) men. Of the 32 participants, 11 (34%) participants had an education level of degree or above. Most participants (21/32, 66%) identified as White (British or Irish). Of the remaining participants, most identified as of South Asian origin (6/32, 19%). Although all participants (32/32, 100%) spoke English, 25% (8/32) of them spoke a different language at home.

Regarding the person the participants cared for, the mean age was 71.4 (SD 23.15) years. It is noteworthy that this distribution was bimodal with a small number of young people receiving care (aged <38 years; 6/32, 19%) and a large number of older people receiving care (aged >60 years; 24/32, 75%). More than half of those receiving care were classified as women (18/32, 56%). Among the 32 participants, there were 8 (25%) carers looking after a spouse or partner, 11 (34%) looking after a parent, and 11 (34%) looking after a son or daughter. Just more than half of them (17/32, 53%) lived with the person they were caring for.

The most common condition or disability of the person receiving care was depression (11/32, 34%), followed by rheumatoid arthritis (10/32, 31%) and osteoarthritis (9/32, 28%). Importantly, many participants reported caring for individuals with multiple conditions and disabilities (22/32, 69%).

At baseline, of the 32 participants, only 5 (16%) participants reported using a health app before, and 8 (25%) reported participating in health education training in relation to caregiving (including diabetes care, parenting for autism, and moving and handling).

In total, 66% (21/32) of the carers were recruited to the subgroup participating in the "think aloud" and semistructured interview component of the study, including 67% (14/21) women and 70% (7/10) men.

**Table 2.** Characteristics of participants and care recipients.

Characteristics	Values
<b>Participants (N=32)</b>	
<b>Sex, n (%)</b>	
Male	10 (31)
Female	22 (69)
Age (years), mean (SD)	57.9 (11.15)
<b>Education level, n (%)</b>	
No formal education	5 (16)
Other	2 (6)
CSE <sup>a</sup> or GCSE <sup>b</sup> or O-level or City and Guilds or NVQ <sup>c</sup> levels 1-2	6 (19)
A-level or higher national diploma or NVQ level 3 or diploma	8 (25)
Degree or equivalent	6 (19)
Higher degree or postgraduate qualification	5 (16)
<b>Ethnicity, n (%)</b>	
British or Irish	21 (66)
Asian or British Asian (Indian or Bangladeshi)	6 (19)
Black or Black British	2 (6)
Other	3 (9)
<b>Relationship with care recipient, n (%)</b>	
Spouse or partner	8 (25)
Daughter or son	11 (34)
Parent (mother, father, mother-in-law, father-in-law, or grandparent)	11 (34)
Other	2 (6)
<b>Live with care recipient, n (%)</b>	
Yes	17 (53)
No	15 (47)
<b>Care recipients (N=32)</b>	
<b>Sex, n (%)</b>	
Male	12 (38)
Female	18 (56)
Other	2 (6)
Age (years), mean (SD)	71.4 (23.15)
<b>Condition or disability, n (%)</b>	
Depression	11 (34)
Rheumatoid arthritis	10 (31)
Osteoarthritis	9 (28)
Respiratory conditions	8 (25)
Diabetes	8 (25)
Dementia	8 (25)
Learning disabilities	5 (16)
Gastrointestinal conditions	4 (13)
Cancer	4 (13)
Visual problems	4 (13)

Characteristics	Values
Cardiac conditions	3 (9)
Multiple sclerosis	2 (6)
Stroke	2 (6)
Other	8 (25)

<sup>a</sup>CSE: Certificate of Secondary Education.

<sup>b</sup>GCSE: General Certificate of Secondary Education.

<sup>c</sup>NVQ: National Vocational Qualification.

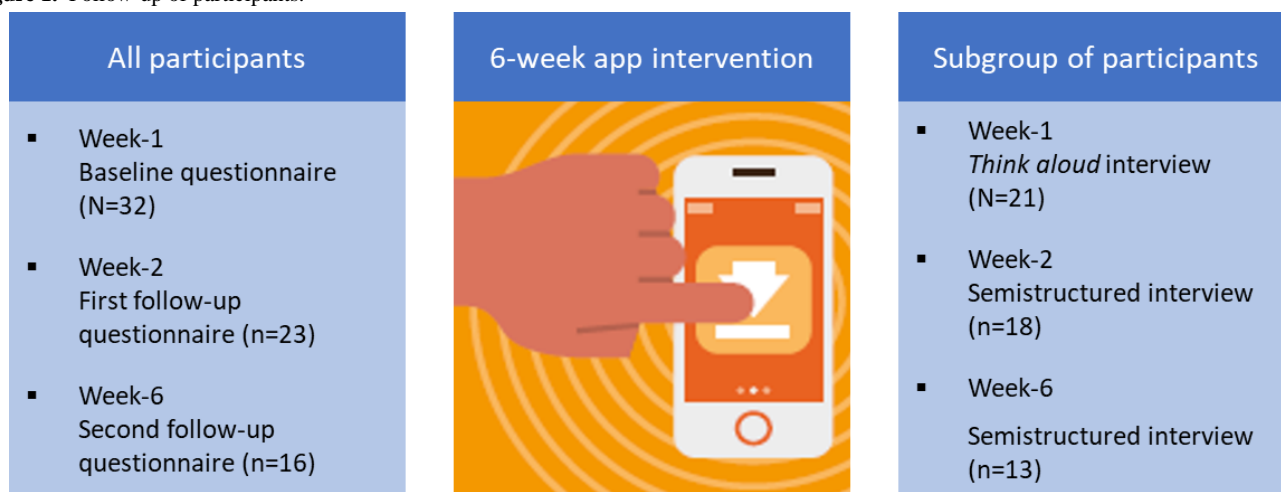
### Loss to Follow-up

All the participants (32/32, 100%) completed the week 1 (baseline) questionnaire. Of the 32 participants, 23 (72%) participants completed the week 2 (first follow-up) questionnaire and 16 (50%) completed the week 6 (second follow-up) questionnaire, with data available for 13 (41%) carers at all 3 time points. There were no significant predictors of withdrawal (carer characteristics, care recipient characteristics, pressure ulcer knowledge, or participant-reported outcome measures) from the study at the  $P < .01$  level.

Comparisons of pressure ulcer knowledge, CSES, PfCS, and CSI were between the baseline and second follow-up measure. The linear mixed models analyses ensured that all available data were used for analyses across time points.

In terms of the qualitative subgroup, 66% (21/32) of the carers participated in the “think aloud” interview, 18 (86%) of whom went on to participate in a semistructured interview at the end of week 2 and 13 (62%) of whom also participated in the semistructured interview at the end of week 6. [Figure 1](#) shows the follow-up of participants across the study.

**Figure 1.** Follow-up of participants.



### Acceptability and Usability of the iCare App

At first follow-up, participants completed SUS and MARS to provide feedback about using the iCare app. [Table 3](#) shows the mean scale scores on these measures. SUS showed that the app has above average usability score (mean 69.94, SD 18.108). The MARS scores supported this finding and showed the app’s overall mean score (mean 3.62, SD 0.540) as being above the midpoint, with the subscales indicating that this is mainly owing to the information-conveying capabilities of the app (mean 4.13, SD 0.452) and its functionality (mean 3.84, SD 0.704). However, the engagement score (mean 3.03, SD 0.669) was midrange, and the perceived impact score (mean 2.01, SD 0.936) was relatively low based on the scale ranges.

The theme, “When you’re not used to these things, they take time to get the hang of;” captured participants’ perceptions and experiences relating to usability. Few participants felt proficient in the use of modern technology at baseline, with many participants describing using only the basic features of their smartphones. Varying levels of familiarity with smartphones and apps in particular were reflected in the amount of time the participants took to find and install the iCare app from the iPhone App Store or the Google Play Store, with downloading times ranging from 1 minute 22 seconds to 15 minutes 34 seconds (average 4 minutes 52 seconds).

Regarding the study’s aim, two key themes were identified from the qualitative data: (1) “When you’re not used to these things, they take time to get the hang of” and (2) “It’s not a fun app but it is informative.” The key themes, subthemes, and illustrative quotations are shown in [Table 4](#).

Once the app was downloaded, some participants found navigating the content more intuitive than others. Some barriers to navigation were related to the design of the app (eg, the indistinguishable nature of pictures on the direct links [tiles] to set modules in the app), whereas others related to relatively common computing functions such as vertical and horizontal scrolling and screen orientation settings. These functions were not considered to be simple or obvious by those participants

who were new to smartphone apps and were identified as a source of frustration in the “think aloud” interviews. Despite these frustrations, many participants thought that they would, with time, learn how to use the app. At the end of the second week, most participants recounted that they had become proficient in the use of the app, which reflects the high usability scores reported using the SUS in the first follow-up questionnaire. However, there were exceptions, including 2 participants who had forgotten their passwords and who had been unable to reset them again. An area that remained as a concern across the 6 weeks was the size of the font on the app. Therefore, some participants expressed a preference for printed forms of information or suggested that a desktop version of the app be made available. However, others recognized that having pressure ulcer information in an app on their mobile devices ensured that advice and support were always available.

The second theme, “It’s not fun but it is informative,” captured participants’ perceptions of and experiences with the performance of the app in terms of conveying information and its functionality and engagement. Participants who were familiar with digital technologies highlighted a missed opportunity by creators to generate an experience beyond the content itself and drew attention to the advantages of app-to-app linking and game mechanisms, which were missing from the iCare app. These participants felt that the addition of these features would have increased their engagement with the app. In contrast, some participants were irritated by animated features (such as the use

of flip cards) because they required more user effort. Although participants disagreed about whether the app should be more entertaining, several participants agreed that great customization and more personalized content would have increased engagement. Regarding customization, participants suggested the addition of bookmarking and favoriting tags, which would have allowed them to return quickly to preferred content, and an activity tracker, which would have tracked their progress. Regarding personalization, a participant suggested the addition of an algorithm that will generate content that is relevant to each user’s personal circumstances. These findings may explain the reasons for the midrange scores for engagement on MARS in the first follow-up questionnaire.

Despite these limitations, participants were united in their description of the app as one that was informative. The highlight for many were the videos, in which the presenter was commended for her pace and use of plain English and for providing a welcome break from the written content. Regardless of baseline levels of knowledge, all participants (32/32, 100%) liked the microlearning approach and endorsed it across the 6-week period. These findings contextualize the high information scores on MARS in the first follow-up interview. Participants described having to juggle their caring responsibilities alongside other responsibilities and that the time they had available to dedicate to learning about pressure ulcers was limited.

**Table 3.** App-related scale scores at the first follow-up.

Scales and parameters	Scores, mean (SD)
System Usability Scale (score range 0-100; higher scores indicate greater usability)	69.94 (18.108)
<b>Mobile App Rating Scale (score range 1-5; higher scores indicate better rating)</b>	
Engagement	3.03 (0.669)
Functionality	3.84 (0.704)
Esthetics	3.46 (0.876)
Information	4.13 (0.452)
Subjective quality score	3.17 (0.978)
Perceived impact of the app	2.01 (0.936)
Quality of the app	3.62 (0.540)

**Table 4.** Themes, subthemes, and illustrative quotes (acceptability and usability of the iCare app).

Themes and subthemes	Quotes
<b>1. When you are not used to these things, they take time to get the hang of</b>	
Limited proficiency in the use of modern technology	<ul style="list-style-type: none"> <li>• “I don’t use mobile phones very often, only for rings.” [think aloud; P104]</li> <li>• “I don’t really do apps...I just use the phone to check my, have I got a text message.” [think aloud; P108]</li> <li>• “I don’t often use [the phone] as a web, for webbing.” [think aloud; P114]</li> <li>• “I’m not very digitally minded...literally, I brought my iPhone to take pictures.” [think aloud; P105]</li> </ul>
Learning over time	<ul style="list-style-type: none"> <li>• “I went to school in the 70s and university in the 80s, so this is not my kind of thing but I could adopt it, I could try.” [think aloud; P116]</li> <li>• “The more I use it, I’ll get the hand of it better.” [think aloud; P101]</li> <li>• “Many people love technology, I love it, I’m crap at it but I want to try and learn.” [think aloud; P110]</li> <li>• “I find it okay because I’ve got used to it now. I’ve looked at it a few times and then I get used to it.” [week 2; P114]</li> <li>• “It seems pretty easy to navigate once you know how it works...Once you learn how to use it, it’s pretty intuitive.” [week 6; P117]</li> </ul>
Small font size	<ul style="list-style-type: none"> <li>• “I have to wear glasses to read and I get tired eyes, watery. I think the writing is, should be a bit bolder.” [think aloud; P108]</li> <li>• “I think I have already said it before, I think the wording needs to be a bit bigger and bolder.” [week 2; P109]</li> </ul>
<b>2. It’s not fun, but it is informative</b>	
Creating an experience beyond the content itself	<ul style="list-style-type: none"> <li>• “I think perhaps you could consider other things like linking it to other systems. For example, like Patient Access...It’s like a GP [general practitioner] practice app where patients can log in, book appointments, repeat prescriptions and things like that. Perhaps you could link it to that because on there, there’s information and support for carers as well.” [think aloud; P106]</li> <li>• “There’s nothing about connecting with...other carers. Nothing about having a discussion about something that you’ve just seen...You could gamify this, that would be more fun...because we spend the entire time reading which is, I get fed up with...being told ‘read this’...I don’t have the time, the energy or the capacity.” [think aloud; P121]</li> <li>• “I suppose they are trying to make it a little bit more interesting, but they could have just done it as bullet points.” [think aloud; P104]</li> <li>• “Because it is not a game. I don’t see [the point] of an extra click. And it makes me feel like I am doing an exam, a multiple-choice exam and it doesn’t make me feel like this is something [I’m going to want to do], I think I would get bored of it.” [think aloud; P111]</li> </ul>
Customization and more personalized content	<ul style="list-style-type: none"> <li>• “I think it would be useful to have some favorites, so sections that you know you’d want to go back to more easily.” [think aloud; P106]</li> <li>• “There isn’t any [customization]. It’s led by the app. It’s just a whole bunch of lines. I can’t customize anything... There isn’t anything that says, I’ve done this bit, and these are the bits that are next.” [think aloud; P121]</li> <li>• “I don’t need that [information] at the moment but if it [was] relevant to my situation... What do they call them now... flow chart! Now that would be useful...so you’re going down a tree until you hit the specific point that you are looking for...I think you have to try and tailor these things.” [think aloud; P113]</li> </ul>
Good use of videos	<ul style="list-style-type: none"> <li>• “I quite like the video content... You don’t want to just read loads and loads and loads of information.” [think aloud; P106]</li> <li>• “I like the way she is talking...a good pace and she was very clear in describing what to look out for...the language she used – it wasn’t really hard terminology.” [think aloud; P107]</li> </ul>
Information is short and to the point	<ul style="list-style-type: none"> <li>• “Everything seems just short and to the point to keep me engaged because, especially as the care you just, your concentration level is just, you’ve just got to be on it, you’re doing other things and also tired...I just need something to spark a little something in me and be simple.” [think aloud; P121]</li> <li>• “It’s not really a fun topic, but it’s very interesting...It was just concise information that someone in my position would need to know, it wasn’t [over the top] with lots of unnecessary information. It was just enough so that I know what to look for and what to do.” [week 2; P118]</li> <li>• “I think it’s very quick and straight to the point most of the time...it’s a very good introduction.” [week 6; P119]</li> </ul>

## Impact of the iCare App on Carers' Knowledge and Confidence in Their Caring Role

Mean scores for pressure ulcer knowledge and participant-reported outcome measures (ie, CSES, PfCS, and CSI) at baseline and 6 week follow-up are reported in [Table 5](#). Regarding the Pressure Ulcer Knowledge Assessment, at baseline, participants had a relatively low score on a scale of 0 to 100 (mean 37.5, SD 16.55). Items on the questionnaire were clustered around four key themes: (1) support surfaces, (2) nutrition and hydration, (3) keep moving, and (4) skin care and inspection. Knowledge—albeit limited—was mainly based on nutrition and hydration (mean 46.8, SD 26.7) and keep moving (mean 37.50, SD 25.40) scale scores. There were deficits in skin care and inspection knowledge (mean 25.63, SD 27.47) and support surfaces knowledge (mean 33.04, SD 20.6).

Regarding participant-reported outcome measures, at baseline, PfCS scores (mean 19.59, SD 6.339) indicated that the group was between “somewhat” and “pretty” well prepared for caring, and CSES scores showed that they were, on average, “somewhat” confident in their ability to care (mean 3.10, SD 0.815). Regarding CSI, carers scored a mean of 11.91 (SD 6.428), representing strain “sometimes.”

Overall, pressure ulcer knowledge scores significantly changed ( $F_{1,6,112}=21.624$ ;  $P<.001$ ) from baseline (mean 37.5, SE 2.926) to the second follow-up (mean 59.72, SE 3.985). From the subscale scores, this difference was likely owing to changes in the “support surfaces” knowledge category, which increased from baseline scores of a relatively low mean of 33.04 (SE 3.653) to a relatively high score (mean 71.11, SE 3.906). Trends toward significant increases in knowledge were found from the subscales for “nutrition and hydration,” “keep moving,” and “skin care and inspection.”

There were no significant results pertaining to participant-reported outcomes across the 6-week period; the PfCS (baseline: mean 27.59, SE 1.121; second follow-up: mean

28.11, SE 1.110;  $P=.60$ ), CSES (baseline: mean 3.10, SE 0.144; second follow-up: mean 3.38, SE 0.172;  $P=.12$ ), and CSI (baseline: mean 11.91, SE 1.136; second follow-up: mean 12.85, SE 1.425;  $P=.47$ ) did not show any significant change.

Regarding this study's aim, only 1 key theme was identified from the qualitative data: “I'm more careful now and would react to signs of redness.” This theme describes participants' reflections on the new knowledge they had acquired, the changes they had made to their caring routines as a result of this new knowledge, their increased vigilance for signs of skin damage, and their intentions regarding the app going forward. Related subthemes and illustrative quotations are shown in [Textbox 1](#).

Participants reported acquiring new knowledge as they progressed through the different modules. Before using the app, some participants had only a rudimentary understanding of the factors contributing to pressure ulcer development, as shown in the baseline pressure ulcer knowledge scores, and had not considered the person they cared for to be especially vulnerable to pressure ulceration because they were neither wheelchair users nor confined to bed. At follow-up, several participants described how they had changed their caring routines as a result of this new knowledge, particularly routines related to movement, patient positioning, and moving and handling. However, most participants felt that the person they cared for was not at high risk of pressure ulceration. As such, the primary learning outcome had not been a change in caring behavior, but an improved understanding of the dangers of pressure ulcers and an increased readiness to react to signs of skin damage. Given that most participants felt that the person they cared for was not at high risk of pressure ulceration, at the final follow-up, most of them felt that they had learned enough about the prevention and management of pressure ulcers and did not envisage returning to the app in the immediate future. However, approximately all participants wanted to retain the app in case their circumstances were to change.

**Table 5.** Pressure ulcer knowledge and participant-reported outcome measures.

Tools and parameters	Baseline, mean (SE)	Second follow-up (week 6), mean (SE)	Test of fixed effects of time	
			<i>F</i> ( <i>df</i> )	<i>P</i> value
<b>Pressure Ulcer Knowledge Assessment (score range 0-100)</b>				
Support surfaces	33.04 (3.653)	71.11 (3.906)	50.415 (1,22.457)	<.001
Nutrition and hydration	46.88 (4.538)	58.62 (4.302)	6.122 (1,21.624)	.02
Keep moving	37.50 (4.490)	57.58 (6.857)	7.365 (1,16.949)	.02
Skin care and inspection	25.63 (4.856)	46.90 (7.611)	7.349 (1,17.299)	.02
Total score	37.25 (2.926)	59.72 (3.985)	29.452 (1,17.850)	<.001
Preparedness for Caregiving Scale (total; score range 0-32)	27.59 (1.121)	28.11 (1.110)	0.286 (1,16.640)	.60
Caregiving Self-Efficacy Scale (mean item; score range 1-5)	3.10 (0.144)	3.38 (0.172)	2.716 (1,16.617)	.12
Caregiver Strain Index (total; score range 0-24)	11.91 (1.136)	12.85 (1.425)	0.553 (1,14.602)	.47

**Textbox 1.** Theme, subthemes, and illustrative quotes (impact of the iCare app on carers' knowledge and confidence in their caring role).

<p><b>Theme</b></p> <ul style="list-style-type: none"> <li>I'm more careful now and would react to signs of redness</li> </ul> <p><b>Subtheme 1</b></p> <ul style="list-style-type: none"> <li>Acquiring new knowledge</li> <li>Quotes <ul style="list-style-type: none"> <li>"There are two pictures and I can see with the pictures it's going to the lady's heel, which shows obviously that if you're sitting too long with your feet up on a surface you could develop a pressure ulcer on your heel, which is something I wouldn't have ever thought of...So, I think there is a lot of good information there, already I can see things that I never knew about pressure ulcers, I just thought it was for people in bed." [think aloud; P118]</li> <li>"That was really, really interesting...You mainly assume, I know about people in wheelchairs, people that are bed ridden but I didn't realize that it could also be people like my mum that's not well for a couple of days...So that's really interesting...that's shocking." [think aloud; P104]</li> </ul> </li> </ul> <p><b>Subtheme 2</b></p> <ul style="list-style-type: none"> <li>Changing care routines</li> <li>Quotes <ul style="list-style-type: none"> <li>"I don't pull him up the bed anymore...I'm turning him more..." [week 2; P101]</li> <li>"I've just made sure, I guess that when I put my wife on the bed...I should lift her up and not drag her...I'm [also] looking...I keep an eye on [her skin]. And one district nurse, some time ago, gave me a [skin barrier] spray [to protect it from excessive moisture], which I spray. But I guess as a result of this video, I'm spraying it more often." [week 2; P109]</li> </ul> </li> </ul> <p><b>Subtheme 3</b></p> <ul style="list-style-type: none"> <li>Alert to signs of redness</li> <li>Quotes <ul style="list-style-type: none"> <li>"Well I certainly know the signs of redness now...so obviously there's a sort of thing to look out for." [week 2; P110]</li> <li>"If my mother did have bedsores...I'd know what to look for...Whereas before I had the app, I wouldn't have had a clue really." [week 2; P112]</li> <li>"I understand when I'm looking at something now better about soreness." [week 2; P114]</li> <li>"The little red, I wouldn't have ever thought of that, if I'd seen a red mark, I would have just thought, oh, wouldn't have thought much of it. But after looking at this, if I ever saw anything like that then that would prompt me to see further help." [week 2; P118]</li> </ul> </li> </ul> <p><b>Subtheme 4</b></p> <ul style="list-style-type: none"> <li>Learned enough for now and keeping the app in case of changing circumstances</li> <li>Quotes <ul style="list-style-type: none"> <li>"There's only so much of it that's relevant to me at the moment. But I know if I need to, like if my godfather for instance gets worse...then I will be able to refer back to it. Yeah, so...if the situation comes up then it's good to have it...I'll definitely keep it on my phone." [week 6; P108]</li> <li>"I don't [use it] as my wife hasn't, at the moment anyway, hasn't got the starting of an ulcer... [But] I'd like to keep it there...I will refer to it from time to time, because it's always a good idea to keep on top of the situation." [week 6; P109]</li> <li>"Because I know it's there if I need it, but like I say, the person I look after, they're quite mobile and I'm quite aware to look out for things...I know it's there so I can go back to it...It's only like if the situation occurred, I might go and double check something." [week 6; P119]</li> </ul> </li> </ul>
---

## Withdrawal Analysis

There were no significant baseline predictors of withdrawal from the study at the  $P < .01$  level. Only pressure ulcer knowledge regarding mobility was associated with withdrawal from the study at the  $P < .05$  level, with an odds ratio of 1.035 (95% CI 1-1.071;  $P = .047$ ). A 10-point increase in this knowledge

increasing the chances of withdrawal by 3.5% (original probability of 16/32, 50%, with related odds of 1).



## Discussion

### Principal Findings

Pressure ulcers are a significant source of burden to informal carers [11-14]. Smartphone apps offer a promising way to support carers by providing access to information and resources at any time where the person has internet connection [18,19]. To the best of our knowledge, no health care app has been evaluated among carers of people at risk of pressure ulceration. The aim of our study was to explore carer perspectives on the acceptability and usability of a pressure ulcer app and determine whether the app increased carers' knowledge and confidence in their caring role.

Despite wide variability in the ease with which carers were able to download and access the app on first use, we found relatively high levels of usability and acceptability among our sample, which comprised informal carers with and those without previous exposure to health care apps. We did not use SUS and MARS at the second follow-up to limit responder burden (especially as loss to follow-up was a concern), but taking these measures again (in a full trial) will be helpful to determine usability or acceptability following a long period with the app, as users mentioned that it will take time to become familiar with the app in the interviews. The information quality was deemed to be especially useful, and participants demonstrated that it improved their knowledge related to pressure ulcer prevention over the pilot period. Although retention of knowledge over the long term is hard to predict, several participants expressed their intention to retain the app and return to the content if they needed to in the future.

The microlearning approach was positively received by participants, who enjoyed the short focused pieces of content. Participants described having to juggle their caring responsibilities alongside other responsibilities and that the time they had available to dedicate to learning about pressure ulcers was limited. iCare does not represent the first use of the microlearning approach in health care apps for informal carers; for example, it is an approach adopted in a mobile app for supporting dementia relatives [39]. However, this study is potentially the first to report carer perspectives of and experiences with knowledge acquisition and skill development using this approach.

There were no significant results pertaining to changes in participant-reported outcomes across the 6-week period. In its current form, the app is generally didactic and underdeveloped in terms of customization and personalized content, which can include, for example, reminders for tasks and deadlines and live support for carers via groups and chats with other carers and professionals. The inclusion of such features can potentially address outcomes including preparedness for caregiving, caregiver self-efficacy, and caregiver strain. This is supported by Grossman et al [40] who suggested that the integration and interaction of five types of app functions successfully relieved caregiver burden: (1) information and resources; (2) practical problem-solving involving behavioral solutions, medication management, safety, and personal health record tracking; (3) memory aids; (4) family communication, including coordinating

care; calendars for appointments and sharing; medical and emergency contact lists; and ability to share important information, photos, and messages among caregivers and family members; and (5) caregiver support (ie, care for the caregiver).

### Limitations

We met the sample size for a small pilot study to provide some indicative parameters that can be built upon for a large randomized controlled trial. However, most participants felt that the person they cared for was not at high risk of pressure ulcers. It would have been useful to have had the pressure ulcer risk assessment score (eg, scores generated using validated tools such as the Waterlow [41] and Braden [42] scales) for the person receiving care to better understand the context within which informal carers were providing care. Furthermore, the inclusion of a larger proportion of carers of people at higher risk of pressure ulcers may have produced more promising results.

The use of linear mixed models analyses helped with the loss to follow-up by allowing all available data to be used; however, the reasons for dropout need to be investigated further.

### Recommendations

This study has demonstrated that microlearning (presenting bite sized chunks of information) is acceptable and useful for users; thus, it is a strategy that can be pursued in further studies and apps of this nature, especially with carer users. However, this must be done with careful thought to the accessibility of mobile apps among the wider population. Not having either a suitable device or an internet connection is the first barrier to the usefulness of mobile apps and may exacerbate inequalities and care. Solutions may include provision of devices and training on downloading and using apps for those first-time users who require additional support, the costs of which will need to be taken into account in any wide-scale roll out. There are also some limitations in the current implementation of the iCare app, which, if addressed, can improve its usability and usefulness. Increasing interactive elements, gamification (potentially using evidence-based behavior change techniques), and customization based on user preferences are potential alterations that can achieve better results on carer-based outcomes.

### Conclusions

This study provides insight into the perceptions of carers on the acceptability of the iCare app and the impact of the app on their pressure ulcer knowledge and confidence in their caring role. The mixed methods analysis found that the app was acceptable to most participants, who endorsed the microlearning approach and perceived the app to be highly informative. In addition, at the end of 6 weeks, carers demonstrated a significant increase in their pressure ulcer knowledge. However, there were no significant results pertaining to participant-reported outcomes. The findings may reflect the need for upcoming iterations of the iCare app to use more interactive elements and introduction of gamification and customization based on user preferences. Future studies will need to capture the risk assessment scores for the person receiving care and sample a broad range of informal carers, including carers of people at high risk of pressure ulceration.

## Acknowledgments

The authors would like to thank UCL Partners who funded this pilot study. The authors would also like to thank all the informal carers who took time to participate in this study. The authors are grateful for the participants' insightful contributions, which have made this study both informative and meaningful.

## Conflicts of Interest

JS is a board member of Care City, the organization that implemented the app. She is also an employee of UCL Partners, the organization that funded the evaluation. All the other authors declare they have no conflicts of interest related to the paper.

## References

1. Pressure ulcer core curriculum. National Health Service England. 2018. URL: <https://www.england.nhs.uk/pressure-ulcer-core-curriculum/> [accessed 2019-08-15]
2. Jaul E, Barron J, Rosenzweig JP, Menczel J. An overview of co-morbidities and the development of pressure ulcers among older adults. *BMC Geriatr* 2018 Dec 11;18(1):305 [FREE Full text] [doi: [10.1186/s12877-018-0997-7](https://doi.org/10.1186/s12877-018-0997-7)] [Medline: [30537947](https://pubmed.ncbi.nlm.nih.gov/30537947/)]
3. Chen HL, Cai JY, Du L, Shen HW, Yu HR, Song YP, et al. Incidence of pressure injury in individuals with spinal cord injury: a systematic review and meta-analysis. *J Wound Ostomy Continence Nurs* 2020;47(3):215-223. [doi: [10.1097/WON.0000000000000633](https://doi.org/10.1097/WON.0000000000000633)] [Medline: [32384524](https://pubmed.ncbi.nlm.nih.gov/32384524/)]
4. Stevenson R, Collinson M, Henderson V, Wilson L, Dealey C, McGinnis E, et al. The prevalence of pressure ulcers in community settings: an observational study. *Int J Nurs Stud* 2013 Nov;50(11):1550-1557. [doi: [10.1016/j.ijnurstu.2013.04.001](https://doi.org/10.1016/j.ijnurstu.2013.04.001)] [Medline: [23764144](https://pubmed.ncbi.nlm.nih.gov/23764144/)]
5. McGinnis E, Briggs M, Collinson M, Wilson L, Dealey C, Brown J, et al. Pressure ulcer related pain in community populations: a prevalence survey. *BMC Nurs* 2014 Jun 21;13:16 [FREE Full text] [doi: [10.1186/1472-6955-13-16](https://doi.org/10.1186/1472-6955-13-16)] [Medline: [25024642](https://pubmed.ncbi.nlm.nih.gov/25024642/)]
6. Gorecki C, Closs SJ, Nixon J, Briggs M. Patient-reported pressure ulcer pain: a mixed-methods systematic review. *J Pain Symptom Manage* 2011 Sep;42(3):443-459 [FREE Full text] [doi: [10.1016/j.jpainsymman.2010.11.016](https://doi.org/10.1016/j.jpainsymman.2010.11.016)] [Medline: [21444182](https://pubmed.ncbi.nlm.nih.gov/21444182/)]
7. Guest JF, Fuller GW, Vowden P. Cohort study evaluating the burden of wounds to the UK's National Health Service in 2017/2018: update from 2012/2013. *BMJ Open* 2020 Dec 22;10(12):e045253 [FREE Full text] [doi: [10.1136/bmjopen-2020-045253](https://doi.org/10.1136/bmjopen-2020-045253)] [Medline: [33371051](https://pubmed.ncbi.nlm.nih.gov/33371051/)]
8. National Stop the Pressure programme: one year on. National Health Service England. URL: <https://www.england.nhs.uk/national-stop-the-pressure-programme-one-year-on/> [accessed 2019-08-15]
9. Guest JF, Fuller GW, Vowden P, Vowden KR. Cohort study evaluating pressure ulcer management in clinical practice in the UK following initial presentation in the community: costs and outcomes. *BMJ Open* 2018 Jul 25;8(7):e021769 [FREE Full text] [doi: [10.1136/bmjopen-2018-021769](https://doi.org/10.1136/bmjopen-2018-021769)] [Medline: [30049697](https://pubmed.ncbi.nlm.nih.gov/30049697/)]
10. Taylor C, Mulligan K, McGraw C. Barriers and enablers to the implementation of evidence-based practice in pressure ulcer prevention and management in an integrated community care setting: a qualitative study informed by the theoretical domains framework. *Health Soc Care Community* 2021 May;29(3):766-779. [doi: [10.1111/hsc.13322](https://doi.org/10.1111/hsc.13322)] [Medline: [33662173](https://pubmed.ncbi.nlm.nih.gov/33662173/)]
11. Nunes JB, Busanello J, Mello-Carpes PB, Cardoso LS, Costa VZ, Deus LM. Conception of health caregivers of individuals with pressure ulcer. *Cienc Cuid Saúde* 2016 May 26;14(4):1462-1470. [doi: [10.4025/ciencucuidsaude.v14i4.25753](https://doi.org/10.4025/ciencucuidsaude.v14i4.25753)]
12. Rodrigues AM, Ferré-Grau C, Ferreira PL. Being an informal caregiver of a person with a pressure ulcer in the Azores Islands. *Adv Skin Wound Care* 2015 Oct;28(10):452-459. [doi: [10.1097/01.ASW.0000471191.11548.dd](https://doi.org/10.1097/01.ASW.0000471191.11548.dd)] [Medline: [26375948](https://pubmed.ncbi.nlm.nih.gov/26375948/)]
13. Rodrigues AM, Ferreira PL, Ferré-Grau C. Providing informal home care for pressure ulcer patients: how it affects carers' quality of life and burden. *J Clin Nurs* 2016 Oct;25(19-20):3026-3035. [doi: [10.1111/jocn.13356](https://doi.org/10.1111/jocn.13356)] [Medline: [27577533](https://pubmed.ncbi.nlm.nih.gov/27577533/)]
14. García-Sánchez FJ, Martínez-Vizcaíno V, Rodríguez-Martín B. Conceptualisations on home care for pressure ulcers in Spain: perspectives of patients and their caregivers. *Scand J Caring Sci* 2019 Sep;33(3):592-599. [doi: [10.1111/scs.12652](https://doi.org/10.1111/scs.12652)] [Medline: [30735262](https://pubmed.ncbi.nlm.nih.gov/30735262/)]
15. NHS England's Commitment to Carers. National Health Service England. 2014 May. URL: <https://www.england.nhs.uk/wp-content/uploads/2014/05/commitment-to-carers-may14.pdf> [accessed 2021-04-27]
16. Sherifali D, Ali MU, Ploeg J, Markle-Reid M, Valaitis R, Bartholomew A, et al. Impact of Internet-based interventions on caregiver mental health: systematic review and meta-analysis. *J Med Internet Res* 2018 Jul 03;20(7):e10668 [FREE Full text] [doi: [10.2196/10668](https://doi.org/10.2196/10668)] [Medline: [29970358](https://pubmed.ncbi.nlm.nih.gov/29970358/)]
17. Ploeg J, Ali MU, Markle-Reid M, Valaitis R, Bartholomew A, Fitzpatrick-Lewis D, et al. Caregiver-focused, Web-based interventions: systematic review and meta-analysis (part 2). *J Med Internet Res* 2018 Oct 26;20(10):e11247 [FREE Full text] [doi: [10.2196/11247](https://doi.org/10.2196/11247)] [Medline: [30368439](https://pubmed.ncbi.nlm.nih.gov/30368439/)]

18. Heynsbergh N, Heckel L, Botti M, Livingston PM. A smartphone app to support carers of people living with cancer: a feasibility and usability study. *JMIR Cancer* 2019 Jan 31;5(1):e11779 [FREE Full text] [doi: [10.2196/11779](https://doi.org/10.2196/11779)] [Medline: [30702432](https://pubmed.ncbi.nlm.nih.gov/30702432/)]
19. Brown EL, Ruggiano N, Page TF, Roberts L, Hristidis V, Whiteman KL, et al. CareHeroes Web and Android™ apps for dementia caregivers: a feasibility study. *Res Gerontol Nurs* 2016;9(4):193-203 [FREE Full text] [doi: [10.3928/19404921-20160229-02](https://doi.org/10.3928/19404921-20160229-02)] [Medline: [29977440](https://pubmed.ncbi.nlm.nih.gov/29977440/)]
20. O' Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. *J Health Serv Res Policy* 2008 Apr;13(2):92-98. [doi: [10.1258/jhsrp.2007.007074](https://doi.org/10.1258/jhsrp.2007.007074)] [Medline: [18416914](https://pubmed.ncbi.nlm.nih.gov/18416914/)]
21. Major A, Calandrino T. Beyond chunking: micro-learning secrets for effective online design. *FDLA J* 2018;3:13 [FREE Full text]
22. Pressure ulcer education pack/booklet. North East London Foundation Trust (NELFT). Essex, UK: NELFT; 2018. URL: <https://www.nelft.nhs.uk/download.cfm?doc=docm93jjm4n3742.pdf&ver=6196> [accessed 2022-09-08]
23. Browne RH. On the use of a pilot sample for sample size determination. *Stat Med* 1995 Sep 15;14(17):1933-1940. [doi: [10.1002/sim.4780141709](https://doi.org/10.1002/sim.4780141709)] [Medline: [8532986](https://pubmed.ncbi.nlm.nih.gov/8532986/)]
24. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Stat Methods Med Res* 2016 Jun;25(3):1057-1073 [FREE Full text] [doi: [10.1177/0962280215588241](https://doi.org/10.1177/0962280215588241)] [Medline: [26092476](https://pubmed.ncbi.nlm.nih.gov/26092476/)]
25. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant* 2018;52(4):1893-1907 [FREE Full text] [doi: [10.1007/s11135-017-0574-8](https://doi.org/10.1007/s11135-017-0574-8)] [Medline: [29937585](https://pubmed.ncbi.nlm.nih.gov/29937585/)]
26. Wu J, Tombor I, Shahab L, West R. Usability testing of a smoking cessation smartphone application ('SmokeFree Baby'): a think-aloud study with pregnant smokers. *Digit Health* 2017 Apr 12;3:2055207617704273 [FREE Full text] [doi: [10.1177/2055207617704273](https://doi.org/10.1177/2055207617704273)] [Medline: [29942594](https://pubmed.ncbi.nlm.nih.gov/29942594/)]
27. Crane D, Garnett C, Brown J, West R, Michie S. Factors influencing usability of a smartphone app to reduce excessive alcohol consumption: think aloud and interview studies. *Front Public Health* 2017 Apr 3;5:39 [FREE Full text] [doi: [10.3389/fpubh.2017.00039](https://doi.org/10.3389/fpubh.2017.00039)] [Medline: [28421175](https://pubmed.ncbi.nlm.nih.gov/28421175/)]
28. De Meyer D, Verhaeghe S, Van Hecke A, Beeckman D. Knowledge of nurses and nursing assistants about pressure ulcer prevention: a survey in 16 Belgian hospitals using the PUKAT 2.0 tool. *J Tissue Viability* 2019 May;28(2):59-69. [doi: [10.1016/j.jtv.2019.03.002](https://doi.org/10.1016/j.jtv.2019.03.002)] [Medline: [30935739](https://pubmed.ncbi.nlm.nih.gov/30935739/)]
29. Kuhn D, Fulton BR. Efficacy of an educational program for relatives of persons in the early stages of Alzheimer's Disease. *J Gerontol Soc Work* 2004 Aug 18;42(3-4):109-130. [doi: [10.1300/j083v42n03\\_07](https://doi.org/10.1300/j083v42n03_07)]
30. Archbold PG, Stewart BJ, Greenlick MR, Harvath T. Mutuality and preparedness as predictors of caregiver role strain. *Res Nurs Health* 1990 Dec;13(6):375-384. [doi: [10.1002/nur.4770130605](https://doi.org/10.1002/nur.4770130605)] [Medline: [2270302](https://pubmed.ncbi.nlm.nih.gov/2270302/)]
31. Thornton M, Travis SS. Analysis of the reliability of the modified caregiver strain index. *J Gerontol B Psychol Sci Soc Sci* 2003 Mar;58(2):S127-S132. [doi: [10.1093/geronb/58.2.s127](https://doi.org/10.1093/geronb/58.2.s127)] [Medline: [12646602](https://pubmed.ncbi.nlm.nih.gov/12646602/)]
32. Brook J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. *Usability Evaluation in Industry*. London, UK: CRC Press; 1995:189-194.
33. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015 Mar 11;3(1):e27 [FREE Full text] [doi: [10.2196/mhealth.3422](https://doi.org/10.2196/mhealth.3422)] [Medline: [25760773](https://pubmed.ncbi.nlm.nih.gov/25760773/)]
34. Ericsson KA, Simon HA. *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA, USA: MIT Press; 1996.
35. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, editors. *Analyzing Qualitative Data*. London, UK: Routledge; 1994:173-194.
36. 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](https://doi.org/10.1186/1471-2288-13-117)] [Medline: [24047204](https://pubmed.ncbi.nlm.nih.gov/24047204/)]
37. Hirani SP, Rixon L, Beynon M, Cartwright M, Cleanthous S, Selva A, WSD investigators. Quantifying beliefs regarding telehealth: development of the Whole Systems Demonstrator Service User Technology Acceptability Questionnaire. *J Telemed Telecare* 2017 May;23(4):460-469. [doi: [10.1177/1357633X16649531](https://doi.org/10.1177/1357633X16649531)] [Medline: [27224997](https://pubmed.ncbi.nlm.nih.gov/27224997/)]
38. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Serv Res* 2017 Jan 26;17(1):88 [FREE Full text] [doi: [10.1186/s12913-017-2031-8](https://doi.org/10.1186/s12913-017-2031-8)] [Medline: [28126032](https://pubmed.ncbi.nlm.nih.gov/28126032/)]
39. Halbach T, Solheim I, Ytrehus S, Schulz T. A mobile application for supporting dementia relatives: a case study. *Stud Health Technol Inform* 2018;256:839-846. [Medline: [30371449](https://pubmed.ncbi.nlm.nih.gov/30371449/)]
40. Grossman MR, Zak DK, Zelinski EM. Mobile apps for caregivers of older adults: quantitative content analysis. *JMIR Mhealth Uhealth* 2018 Jul 30;6(7):e162 [FREE Full text] [doi: [10.2196/mhealth.9345](https://doi.org/10.2196/mhealth.9345)] [Medline: [30061093](https://pubmed.ncbi.nlm.nih.gov/30061093/)]
41. Waterlow J. Pressure sores: a risk assessment card. *Nurs Times* 1985;81(48):49-55. [Medline: [3853163](https://pubmed.ncbi.nlm.nih.gov/3853163/)]
42. Braden B, Bergstrom N. A conceptual schema for the study of the etiology of pressure sores. *Rehabil Nurs* 1987;12(1):8-12. [doi: [10.1002/j.2048-7940.1987.tb00541.x](https://doi.org/10.1002/j.2048-7940.1987.tb00541.x)] [Medline: [3643620](https://pubmed.ncbi.nlm.nih.gov/3643620/)]

## Abbreviations

**CSES:** Caregiving Self-Efficacy Scale  
**CSI:** Caregiver Strain Index  
**MARS:** Mobile App Rating Scale  
**NHS:** National Health Service  
**PfCS:** Preparedness for Caregiving Scale  
**SUS:** System Usability Scale

*Edited by A Mavragani; submitted 16.01.22; peer-reviewed by A Samuel, C Memering, S Davis; comments to author 12.07.22; revised version received 22.07.22; accepted 23.07.22; published 16.09.22.*

*Please cite as:*

*McKeown E, McGraw C, Holder P, Shand J, Hirani SP*

*Acceptability and Impact of an Educational App (iCare) for Informal Carers Looking After People at Risk of Pressure Ulceration: Mixed Methods Pilot Study*

*JMIR Form Res 2022;6(9):e36517*

*URL: <https://formative.jmir.org/2022/9/e36517>*

*doi: [10.2196/36517](https://doi.org/10.2196/36517)*

*PMID: [36112413](https://pubmed.ncbi.nlm.nih.gov/36112413/)*

©Eamonn McKeown, Caroline McGraw, Pru Holder, Jenny Shand, Shashivadan P Hirani. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 16.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Waiting Lists for Psychotherapy and Provider Attitudes Toward Low-Intensity Treatments as Potential Interventions: Survey Study

Allison Peipert<sup>1</sup>, BS; Anne C Krendl<sup>1</sup>, PhD; Lorenzo Lorenzo-Luaces<sup>1</sup>, PhD

Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, United States

**Corresponding Author:**

Lorenzo Lorenzo-Luaces, PhD  
Department of Psychological and Brain Sciences  
Indiana University Bloomington  
1101 E 10th St  
Bloomington, IN, 47405  
United States  
Phone: 1 812 855 2012  
Email: [llorenz@indiana.edu](mailto:llorenz@indiana.edu)

## Abstract

**Background:** Common mental disorders, including depression and anxiety, are leading causes of disability worldwide. Digital mental health interventions, such as web-based self-help and other low-intensity treatments (LITs) that are not digital (eg, bibliotherapy), have the potential to reach many individuals by circumventing common barriers present in traditional mental health care. It is unclear how often LITs are used in clinical practice, or whether providers would be interested in their use for treatment waiting lists.

**Objective:** The aims of this study were to (1) describe current practices for treatment waiting lists, (2) describe providers' attitudes toward digital and nondigital LITs for patients on a waiting list, and (3) explore providers' willingness to use digital and nondigital LITs and their decisions to learn about them.

**Methods:** We surveyed 141 practicing mental health care providers (eg, therapists and psychologists) and provided an opportunity for them to learn about LITs.

**Results:** Most participants reported keeping a waiting list. Few participants reported currently recommending digital or nondigital LITs, though most were willing to use at least one for patients on their waiting list. Attitudes toward digital and nondigital LITs were neutral to positive. Guided digital and nondigital LITs were generally perceived to be more effective but less accessible, and unguided interventions were perceived to be less effective but more accessible. Most participants selected to access additional information on LITs, with the most popular being web-based self-help.

**Conclusions:** Results suggest providers are currently not recommending LITs for patients on treatment waiting lists but would be willing to recommend them. Future work should explore barriers and facilitators to implementing digital and nondigital LITs for patients on treatment waiting lists.

(*JMIR Form Res* 2022;6(9):e39787) doi:[10.2196/39787](https://doi.org/10.2196/39787)

**KEYWORDS**

psychotherapy; CBT; cognitive behavioral therapy; behavior therapy; digital mental health; self-help; support group; mental health; digital health; eHealth; low-intensity intervention; survey; waiting list; health system; health care delivery; health care professional; care provider; bibliotherapy; attitude; perception; digital intervention; web-based intervention; depression; anxiety; mental disorder

## Introduction

Common mental disorders (CMDs), such as depression, anxiety, and insomnia, are leading causes of disability worldwide [1-3] and a significant burden to the health care system [4]. Mental health needs have increased with the COVID-19 pandemic,

such that approximately 3 in 10 adults in the United States report clinically elevated anxiety or depression symptoms—a number that may have tripled since 2019 [5]. Effective treatments exist for CMDs, including psychotherapy and pharmacotherapy. However, despite the prevalence of CMDs and the efficacy of existing treatments, few people with mental health concerns receive any form of treatment [6,7].

Barriers related to receiving traditional psychotherapy or other forms of treatment include time availability, lack of financial resources (including insurance coverage), stigma, and low provider availability [6,8]. During the COVID-19 pandemic, many of these barriers have been amplified, particularly provider availability and waiting time for treatment. Increased delays in treatment have been reported across a number of health service areas, including oncology, elective surgery, and general health care [9,10]. Waiting lists are used by providers as the demand for services often surpasses provider capacity. Generally, this involves individuals seeking treatment to be placed on a list prior to receiving a scheduled appointment or scheduling appointments far in the future. Existing work suggests that treatment waiting lists are commonly used for mental health services, but there is substantial variability in their design and management [11]. A relatively agreed-upon feature of waiting lists is that they are associated with a rather long time waiting for psychological services, raising a number of ethical issues, and forcing individuals to seek out alternate services.

Waiting to receive treatment for mental health specifically can have detrimental effects. For example, waiting for treatment can be associated with increased symptom severity, including symptom deterioration [12]. Additionally, a greater waiting time is associated with a lower likelihood of ultimately engaging with the treatment [13]. Finally, the pretreatment waiting times may also be associated to worse patient engagement even once patients initiate treatment [14,15], decreased probability of improvement with treatment, and increased risk of dropout [13,16]. Patients themselves also identify waiting for treatment as a barrier to care, with some studies reporting negative psychological and behavioral outcomes when placed on treatment waiting lists for mental health care [17].

Current efforts to address barriers to mental health treatment have made little impact on the burden of mental health. For example, research in places where the number of mental health providers has increased show public health burdens of mental illness remain unchanged [18]. Furthermore, some researchers report the public health burden of CMDs, such as mood and anxiety disorders, would show little reduction even if the number of providers doubled instantly [6]. These findings suggest that the current model of mental health care needs to change in order to accommodate the burden of CMDs [7].

Various low-intensity treatments (LITs) exist that are relatively scalable and have demonstrated efficacy for depression, anxiety, and other CMDs. Many of these LITs can be used with additional guidance by a professional or paraprofessional (ie, guided) or can be self-guided by the individual user (ie, unguided) and can be accessed through a digital or nondigital platform. Although face-to-face therapy is one of the most well-studied treatments for CMDs, both guided and unguided LITs have been proven to be effective relative to controls like waiting lists and care as usual [19-22]. In general, unguided self-help is better than control groups (eg, sham internet applications and care as usual), and guided self-help is more efficacious than both controls and unguided self-help [23], with guided self-help appearing to have similar efficacy to face-to-face therapy [24]. One commonly studied nondigital LIT is bibliotherapy, a form of self-help using print materials

[25]. Meta-analyses of randomized controlled trials of bibliotherapy for depression support its efficacy, yielding large mean effect sizes [19,20]. Guided and unguided digital mental health interventions (DMHIs), including internet-based cognitive behavioral therapy (iCBT) or mental health apps, are effective at treating depression relative to controls like waiting lists and care as usual [24,26]. Together, these studies suggest that LITs provide a tenable and effective treatment alternative for people with CMDs.

Despite research supporting the efficacy of digital and nondigital LITs, it is unclear to what extent they are used in clinical practice [27]. We propose that one place where digital and nondigital LITs may be impactful is in the provision of services to individuals waiting for treatment [28]. The period in which people are waiting for psychological services is an important time for intervention because individuals have already overcome some barriers to treatment seeking, including stigma and the lack of contact with a mental health provider, only to be faced with another barrier (ie, time waiting without receiving services). Schleider et al [28] made a similar observation in a small open trial, where they offered a nondigital LIT to patients on a waiting list. We could not identify a study where provider attitudes toward digital or nondigital LITs for patients on waiting lists were explored.

There is limited research on the use of LITs for patients on waiting lists; however, preliminary studies have supported the feasibility of implementing mental health apps for patients waiting for treatment [29,30]. Other researchers have explored barriers and facilitators of DMHI implementation in clinical practice more broadly. In one study comparing implementers and nonimplementers of iCBT, the two groups differed significantly in their perceived knowledge, confidence in when to recommend iCBT, perception of technical problems, organizational resources, and patient referral process allowing for iCBT inclusion [27]. Thus, there may be differences in the perceptions and attitudes of those who choose to implement versus those who choose not to implement digital and nondigital LITs, such as iCBT, in their practice. However, we know little about general attitudes toward LITs and, specifically, their potential to be implemented in clinical practice for patients on a waiting list. This is important for two reasons. First, extensive work has shown that attitudes predict behavior [31]; thus, negative attitudes toward digital and nondigital LITs may predict reduced willingness to endorse them. Second, identifying the factors that dissociate between willingness and unwillingness to implement digital and nondigital LITs could inform how future work should shape interventions to promote providers' LIT use, particularly for practices with prohibitively long wait lists.

The aims of our study were the following: (1) describing current practices surrounding waiting lists, including how often they are used and what providers do with individuals on their waiting lists (aim 1); (2) describing providers' attitudes toward digital and nondigital LITs for patients on their waiting lists (aim 2); and (3) exploring predictors of providers' willingness to use digital and nondigital LITs and their decisions to learn more about them (aim 3).

## Methods

We surveyed currently practicing mental health care providers on their attitudes toward LITs, including guided and unguided bibliotherapy and DMHIs. Additionally, we included a behavioral task that provided an option for participants to receive additional information about LITs.

### Recruitment and Eligibility

Participants were recruited using a survey link posted via emails to professional organization listserves, specifically the American Psychological Association Division 29—Society for the Advancement of Psychotherapy, the Association for Behavioral and Cognitive Therapies, and social media (ie, Twitter). The survey, which was advertised as a survey on “waiting lists and possible resources,” began with a Study Information Sheet outlining the purpose of the research, eligibility criteria, limits of confidentiality, risks and benefits, and compensation. Participants were eligible to participate if they identified as being (1) over the age of 18 years, (2) a practicing licensed mental health care professional, and (3) currently providing at least 1 hour of clinical services per week. We required participants to have to conduct at least an hour of clinic work to ensure participants were currently providing at least some clinical services, but we did not limit the study to those whose only duties were clinical work.

We received 145 survey responses. Two responses were removed after being determined to be from the same individual. Two responses were removed for not providing responses to all questions. The remaining 141 participants were included in data analysis.

### Ethics Approval

Study procedures were approved by Indiana University Bloomington’s Human Subjects and Institutional Review Board (10503).

### Data

The survey, which can be found in the Open Science Framework website [32], was divided into 4 sections. Section 1 collected demographic information (eg, age, gender, race, and ethnicity) and clinical background (eg, state or country of licensure, level of education, clinical practice setting, psychotherapy theoretical orientation, satisfaction with clinical work, years of clinical experience, and average hours of clinical services provided per week).

Section 2 was designed to address aim 1 of our study. In this section, participants were asked to provide information on their current waiting list practices. This included whether their current clinical setting kept a treatment waiting list (vs not keeping a waiting list or scheduling appointments several months in advance), the estimated time patients spent on a waiting list, the effect of the COVID-19 pandemic on waiting time, and current actions taken for patients on a treatment waiting list.

In section 3, participants were provided with a brief description of the following LITs: unguided bibliotherapy, guided bibliotherapy, unguided web-based self-help, guided web-based self-help, and patient support groups. We chose these LITs

because they have been relatively well researched, and research suggests they are effective. Descriptions of each LIT type were provided to ensure that respondents were equally familiar with each intervention. To address aim 2 of our study, for each LIT, respondents were asked to rate their perceived effectiveness (eg, “I believe this option would be effective for patients on a treatment waiting list”), availability (eg, “I believe this option is available and accessible to use with patients on my waiting list”), and willingness to use LIT (eg, “I am willing to use this modality for patients on a waiting list”). Responses were given on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

Section 4 addressed aim 3 of our study by asking respondents if they would like to receive any additional information about the following LITs: bibliotherapy, web-based self-help, or patient support groups. They were subsequently offered information on each LIT, and we tracked which participants chose to receive more information (ie, whether they engaged in information-seeking behavior).

### Missing Data

Missingness in demographic and clinical variables was relatively low (0%-1.4%), with the exception of age (13.5%). Missingness in the attitude variables was greater (10.6%-12.1%), reflecting survey dropout (ie, participants who answered no subsequent questions). We completed regression analyses with the original data set and an imputed data set. The latter imputed missing values for all variables included in the regression analyses using a machine learning algorithm with random forests using the R package “missForest” [33].

### Statistical Analysis

Analyses were conducted using the R programming language (version 4.2.1; the R Core Team) [34]. To describe waiting list practices, we present response frequencies and descriptive statistics (aim 1). We compared descriptive statistics and response frequencies of providers’ perceived efficacy, availability and accessibility, as well as willingness to use the different LITs (aim 2).

To explore predictors of willingness to use digital and nondigital LITs (aim 3), we ran 5 linear regressions, one for each LIT under consideration. Willingness to use each specific intervention was regressed on demographic information (ie, gender, age, and education), professional background (ie, theoretical orientation and practice setting), and clinical variables (ie, the use of a waiting list, clinical satisfaction, and clinical hours per week). In these models, an average “willingness” response variable was included in the regression to control for participant willingness to use other interventions, excluding the one being predicted. Regressions were completed using the original data set and the imputed data set.

To explore predictors of information seeking for digital and nondigital LITs (aim 3), we conducted a series of regressions to explore demographic, professional background and attitudinal predictors of requesting additional information. Specifically, we conducted 3 binomial logistic regressions with the dependent variables of requesting additional information (ie, selecting information vs not selecting information) on the following: (1)

web-based self-help, (2) bibliotherapy, and (3) patient support groups. Each model included demographic information (eg, gender, education, and age), clinical variables (eg, theoretical orientation, clinical satisfaction, years of experience, clinical hours, and presence of a waiting list), modality-specific attitudes (eg, willingness to use the specific intervention and perceived availability and accessibility of the specific intervention), and controls for overall rating tendency (eg, average willingness to use other interventions and average availability or accessibility of other interventions).

### Transparency and Openness

We report how we determined our sample, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and research materials are available on the Open Science Framework website [32]. This study's design and its

analysis were not preregistered. No other papers currently use these data.

## Results

### Sample Characteristics

Demographic and clinical variables of the sample are summarized in Table 1. The sample primarily included female, non-Hispanic, White participants. Participants' average age was about 39 years. Most participants in the sample had a PhD degree and were prescribed to an orientation (eg, cognitive or third-wave behavioral therapy) related to cognitive behavioral therapy (CBT). About half of the participants were employed in a private practice setting with an average of 10 (SD=10.4) years of clinical experience.



**Table 1.** Demographic and clinical variables of 141 providers who responded to a survey of waiting lists and low-intensity treatments.

Demographics	Values
Age (years), mean (SD)	39.2 (10.1)
<b>Gender, n (%)</b>	
Female	92 (65.2)
Male	45 (31.9)
Nonbinary	2 (1.4)
No answer	2 (1.4)
<b>Race or ethnicity, n (%)</b>	
Non-Hispanic White	115 (81.5)
Non-Hispanic Black	1 (0.7)
Hispanic	7 (5)
Asian	6 (4.3)
AIAN <sup>a</sup> , MENA <sup>b</sup> , NHPI <sup>c</sup> , or other	6 (4.3)
Multiracial	6 (4.3)
<b>Education, n (%)</b>	
Associate of Arts degree	1 (0.7)
Bachelor of Arts degree	3 (2.1)
Master of Arts degree	17 (12.1)
PhD	99 (70.2)
PsyD	18 (12.8)
Other	3 (2.1)
Private practice (vs no private practice), n (%)	71 (50.7)
Clinical orientation—CBT <sup>d</sup> (vs other), n (%)	125 (88.7)
<b>Clinical satisfaction, n (%)</b>	
Not satisfied	0 (0)
Slightly satisfied	7 (5)
Neutral	23 (16.4)
Very satisfied	80 (57.1)
Extremely satisfied	30 (21.4)
Clinical experience (years), mean (SD)	10.4 (10.4)
Clinical hours per week, mean (SD)	18.3 (10.9)

<sup>a</sup>AIAN: American Indian or Alaska Native.

<sup>b</sup>MENA: Middle Eastern or North African.

<sup>c</sup>NHPI: Native Hawaiian or Pacific Islander.

<sup>d</sup>CBT: cognitive behavioral therapy.

### Waiting List Characteristics

The majority of survey respondents (n=94, 69.1%) endorsed keeping a formal treatment waiting list (Table 2) with the estimated waiting time averaging about 13 weeks. Others reported not keeping a waiting list but scheduling patients “a couple of months in advance” (n=23, 16.9%). Most of the respondents who endorsed keeping a waiting list also noted their estimated waiting time had been impacted by the

COVID-19 pandemic (n=68, 73%) and estimated waiting times around 9 weeks prior to the pandemic.

The majority of participants who kept a waiting list or scheduled patients for months in advance endorsed taking names or contact information, completing an unstructured brief assessment, and providing referrals. Few reported completing a structured assessment or provided brief psychoeducation, and even fewer (n=13-20, <20%) reported providing information on apps, books, or support groups.

**Table 2.** Features of treatment waiting lists for 141 providers who responded to our survey on waiting lists and low-intensity treatments.

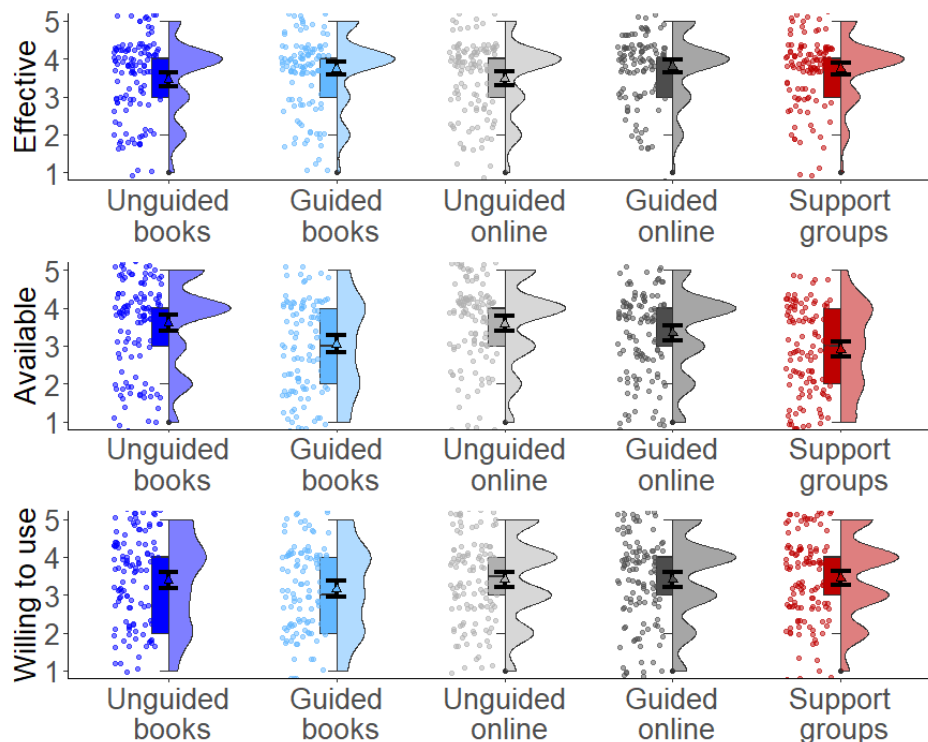
Characteristics	Values	Missing values, n (%)
<b>Waiting list kept, n (%)</b>		5 (3.5)
Yes	94 (69.1)	
No	23 (16.9)	
No, but scheduling appointments months in advance	19 (14)	
<b>Impacted by COVID-19, n (%)</b>		1 (1.1)
Yes	68 (73.1)	
No	25 (26.9)	
<b>Wait time, mean (SD)</b>		
Current wait time (weeks)	12.8 (10.6)	
Pre-COVID-19 wait time (weeks)	9.4 (11.7)	
<b>Waiting list resources used by therapists who kept a waiting list (n=106), n (%)</b>		7 (6.2)
Name or contact information	101 (95)	
Unstructured assessment	77 (72.6)	
Structured assessment	24 (22.6)	
Referrals (psychology or psychiatry)	88 (83)	
Brief psychoeducation	29 (27.4)	
Books	17 (16)	
Apps	13 (12.3)	
Support groups	20 (18.9)	
Other	9 (8.5)	

### Attitudes Toward Digital and Nondigital Low-Intensity Treatments

Figure 1 shows respondents' ratings of willingness to use, perceived efficacy, and availability of different interventions. Average responses were generally between 3 ("neutral") and 4

("somewhat agree"). The exception to this was the perceived low accessibility and availability of patient support groups (mean 2.9, SD 1.2). Average attitudes were highly correlated; specifically, the willingness to use LITs was related to its perceived efficacy ( $r_{122} = 0.76$ , 95% CI 0.67-0.83;  $P < .001$ ).

**Figure 1.** Willingness to use, perceived availability, and perceived efficacy of different low-intensity interventions presented to 141 mental health providers. Box plot indicates median and interquartile range. Triangle and error bars indicate mean values and 95% CI. Violin plot shows distribution of the data.



### Predictors of Willingness to Use an Intervention

We used multiple linear regressions to identify demographic, clinical, waiting list, and attitude variables predicting participants' ratings of willingness to use a LIT (Table S1 in [Multimedia Appendix 1](#)). Across models, averaged willingness to use other LITs predicted the specific likelihood of willingness to use another LIT. For example, the willingness to use unguided bibliotherapy was strongly predicted by the average willingness to use support groups, guided bibliotherapy, and web-based self-help ( $\beta=.65$ , 95% CI 0.45-0.84).

Other predictors of willingness to use a LIT varied across the LITs probed. Compared to individuals in other practice settings, individuals in private practice were more willing to use unguided bibliotherapy ( $\beta=.22$ , 95% CI 0.02-0.42) and less willing to use guided bibliotherapy ( $\beta=-.20$ , 95% CI  $-0.35$  to  $-0.04$ ). Additionally, individuals who reported a CBT theoretical orientation (vs not reporting one) were more willing to use unguided web-based self-help ( $\beta=.22$ , 95% CI 0.08-0.35). There were no consistent significant predictors of willingness to use

support groups. More detailed information is presented in Table S1 in [Multimedia Appendix 1](#).

### Information Seeking About Digital and Nondigital Low-Intensity Treatments

Most ( $n=85$ , 70%) respondents indicated they would like to receive additional information about at least one of these modalities: web-based self-help ( $n=78$ , 63%), bibliotherapy ( $n=66$ , 53%), or support groups ( $n=61$ , 50%). We explored demographic, clinical, waiting list, and attitude variables predicting participants' responses to receive additional information. Overall, there were few strong predictors of information seeking for LITs (Table 3 and Table 4). Willingness to use the specific LIT was predictive of information seeking for web-based self-help (odds ratio [OR]=2.20, 95% CI 1.20-4.28) and bibliotherapy (OR=1.66, 95% CI 1.02-2.80). Relative to individuals in other practice settings, individuals in private practice were more likely to seek information for web-based self-help (OR=5.33, 95% CI 1.48-23.42). There were no consistent predictors of seeking information about support groups.

**Table 3.** Logistic regression analyses predicting providers' information seeking for low-intensity interventions from demographic, clinical, and practice predictors in original data sets.

Original data (N=104)	Web-based self-help			Bibliotherapy			Support groups		
	OR <sup>a</sup> (95% CI)	Z <sup>b</sup>	P value	OR (95% CI)	Z	P value	OR (95% CI)	Z	P value
Intercept	0.16 (0.00-8.08)	-0.91	.36	0.56 (0.01-31.65)	-0.28	.78	0.11 (0.00-5.37)	-1.10	.27
Availability (specific)	0.71 (0.41-1.20)	-1.23	.22	0.94 (0.55-1.58)	-0.23	.82	0.98 (0.66-1.43)	-0.12	.90
Willingness (specific)	2.20 (1.20-4.28)	2.45	.01 <sup>c</sup>	1.66 (1.02-2.80)	1.99	.05 <sup>c</sup>	1.24 (0.77-2.01)	0.87	.38
Availability (general)	1.51 (0.77-3.04)	1.19	.23	1.08 (0.55-2.14)	0.22	.83	0.97 (0.52-1.79)	-0.11	.91
Willingness (general)	1.29 (0.59-2.84)	0.65	.51	1.64 (0.84-3.38)	1.41	.16	1.70 (0.88-3.44)	1.54	.12
Age	0.97 (0.92-1.01)	-1.36	.17	0.99 (0.94-1.04)	-0.46	.64	0.97 (0.93-1.02)	-1.13	.26
Female (vs male)	0.41 (0.12-1.20)	-1.57	.12	0.84 (0.31-2.24)	-0.35	.73	0.69 (0.27-1.75)	-0.78	.44
Doctorate (vs AA/BA/MA <sup>d</sup> )	1.50 (0.30-6.76)	0.52	.60	1.38 (0.33-5.55)	0.45	.65	2.20 (0.56-9.21)	1.12	.26
Keeps a waitlist (vs not)	2.52 (0.64-10.25)	1.33	.18	1.36 (0.41-4.64)	0.50	.62	0.51 (0.14-1.72)	-1.07	.29
Clinical satisfaction (1-5)	0.84 (0.39-1.77)	-0.46	.65	0.56 (0.27-1.10)	-1.64	.10	1.00 (0.51-1.99)	0.01	.99
Clinical hours	1.00 (0.96-1.05)	0.20	.84	1.00 (0.96-1.05)	0.20	.85	0.99 (0.95-1.03)	-0.50	.62
Private practice (vs other)	5.33 (1.48-23.46)	2.40	.01 <sup>c</sup>	2.52 (0.84-8.32)	1.60	.11	2.14 (0.72-6.84)	1.34	.18
CBT <sup>e</sup> (vs other orientation)	0.95 (0.20-4.42)	-0.06	.95	0.41 (0.09-1.74)	-1.18	.24	1.75 (0.41-8.28)	0.74	.46

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

<sup>b</sup>Z statistic from specific model term in logistic regression.

<sup>c</sup>P values are significant when  $P < .05$ .

<sup>d</sup>AA/BA/MA: Associate of Art degree/ Bachelor of Arts degree/Master of Arts degree.

<sup>e</sup>CBT: cognitive behavioral therapy.

**Table 4.** Logistic regression analyses predicting providers' information seeking for low-intensity interventions from demographic, clinical, and practice predictors in imputed data sets.

Imputed data (N=141)	Eb-based self-help			Bibliotherapy			Support groups		
	OR <sup>a</sup> (95% CI)	Z <sup>b</sup>	P value	OR (95% CI)	Z	P value	OR (95% CI)	Z	P value
Intercept	0.05 (0.00-1.23)	-1.79	.07	0.12 (0.00-3.20)	-1.25	.21	0.17 (0.01-3.42)	-1.15	.25
Availability (specific)	0.79 (0.47-1.27)	-0.96	.34	1.01 (0.62-1.61)	0.03	.98	0.96 (0.67-1.36)	-0.25	.8
Willingness (specific)	1.95 (1.17-3.38)	2.48	.01 <sup>c</sup>	1.50 (0.98-2.32)	1.87	.06	1.45 (0.98-2.19)	1.81	.07
Availability (general)	1.74(0.95-3.27)	1.78	.08	1.42 (0.76-2.73)	1.10	.27	1.13 (0.65-1.97)	0.42	.68
Willingness (general)	1.14 (0.59-2.17)	0.40	.69	1.48 (0.81-2.77)	1.27	0.20	1.24 (0.69-2.27)	0.42	.48
Age	0.98 (0.93-1.02)	-1.06	.29	1.00 (0.95-1.04)	-0.15	.88	0.98 (0.94-1.02)	-0.92	.36
Female (vs male)	0.57 (0.21-1.44)	-1.16	.24	1.00 (0.41-2.42)	0.01	.99	0.75 (0.34-1.64)	-0.71	.48
Doctorate (vs AA/BA/MA <sup>d</sup> )	0.79 (0.28-2.14)	-0.46	.65	0.84 (0.33-2.06)	-0.39	.70	0.96 (0.40-2.26)	-0.09	.93
Keeps a waitlist (vs not)	3.32 (1.10-10.30)	2.12	.03 <sup>c</sup>	2.63 (0.94-7.73)	1.81	.07	0.91 (0.33-2.47)	-0.18	.86
Clinical satisfaction (1-5)	1.21 (0.66-2.25)	0.62	.54	0.79 (0.44-1.40)	-0.81	.42	1.18 (0.69-2.01)	0.60	.55
Clinical hours	1.00 (0.96-1.05)	0.14	.89	1.01 (0.97-1.05)	0.28	.78	0.99 (0.95-1.03)	-0.55	.58
Private practice (vs other)	2.85 (1.06-8.42)	2.00	.05 <sup>c</sup>	1.34 (0.55-3.38)	0.64	.52	1.18 (0.50-2.86)	0.37	.71
CBT <sup>d</sup> (vs other orientation)	0.75 (0.19-2.59)	-0.45	.65	0.37 (0.09-1.30)	-1.48	.14	1.07 (0.33-3.39)	0.12	.90

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

<sup>b</sup>Z statistic from specific model term in logistic regression.

<sup>c</sup>P values are significant when  $P < .05$ .

<sup>d</sup>AA/BA/MA: Associate of Art degree/ Bachelor of Arts degree/Master of Arts degree.

<sup>d</sup>CBT: cognitive behavioral therapy.

## Discussion

### Principal Findings

The aims of this study were to describe current waiting list practices (aim 1), describe providers' attitudes toward digital and nondigital LITs for patients on their waiting lists (aim 2), and explore predictors of providers' willingness to use digital and nondigital LITs and their decisions to learn more about them (aim 3). Most providers ( $n=94$ , 69%) endorsed keeping a treatment waiting list. Among those who said they do not, nearly half ( $n=19$ , 45%) reported scheduling patients in the "distant future," for example, 2-3 months away. Thus, most ( $n=113$ , 83%) providers in this sample had an opportunity to use LITs with people waiting for treatment. However, fewer than 20% ( $n=13-20$ ) reported having recommended books, apps, or support groups for patients on a waiting list. The majority of those who endorsed maintaining a waiting list noted the estimated waiting time for their patients to access treatment was high and had increased since the COVID-19 pandemic. The difference in the average estimated waiting time currently and prior to the pandemic was on average about a month, with a significant increase, suggesting patients have experienced an increased delay in accessing mental health care since the onset of the pandemic.

This study has a number of limitations to consider. First, the generalizability of the study results may be limited due to the relatively small sample size and the representativeness of our study sample, which consisted primarily of CBT-oriented providers. However, our sample was relatively similar to the gender and racial or ethnic demographic characteristics of the psychology workforce in the United States [35], and CBT theoretical orientations have become the most popular among providers [36,37]. Additionally, it is possible that providers who are more interested in digital or nondigital LITs may have been more likely to participate in the study, resulting in a biased rating of attitudes relative to the population of mental health providers. Provider estimates of waiting times may likewise be biased or inaccurate in other ways. Previous research on waiting times for psychological services suggests that average waiting time ranges from 2-3 weeks to 2-3 months [38,39], but this varies across studies. Future studies could explore alternative methods, including the "secret shopper" methods, to get more accurate data on waiting lists and estimated waiting times. Lastly, this study assessed providers' limited attitudes toward the listed interventions, specifically their perceptions of the intervention efficacy, accessibility, and their willingness to use it. Future research should investigate more detailed perceptions of these interventions through, for example, qualitative interviews to identify specific barriers and facilitators to digital and nondigital LIT implementation and use.

Despite its limitations, this study has notable strengths. To our knowledge, this is the first survey of providers to assess attitudes toward digital and nondigital LITs for patients on a waiting list. Additionally, beyond assessing attitudes, we also provided an opportunity to learn more about digital and nondigital LITs and assessed participants' decisions to request additional information. Most providers were not currently recommending digital or nondigital LITs for patients on their waiting list, but attitudes toward the interventions were neutral to positive. Most were willing to use at least one intervention for patients on their waiting list. We found no evidence that providers had more positive attitudes regarding digital versus nondigital interventions. Generally, guided interventions were seen as more effective but less accessible than unguided interventions, which, in turn, were seen as less effective but more accessible. Together, these findings support our proposal that dissemination and implementation of digital and nondigital LITs (eg, bibliotherapy or DMHIs) while patients are on waiting lists could be a promising strategy to reduce the burden of untreated CMDs.

Although we found few significant predictors of attitudes and information seeking, we found that practice setting (ie, private practice vs other settings) was a predictor of attitudes and behaviors toward LITs. Individuals in private practice were more willing to use unguided bibliotherapy compared to individuals in other practice settings but were less willing to use guided bibliotherapy and unguided web-based self-help. Interestingly, at the end of our survey, individuals in private practice were more willing to learn about web-based self-help resources compared to individuals in other practice settings. These findings may reflect a knowledge gap wherein individuals in private practice currently do not perceive themselves to know which DMHIs to turn to, a barrier reported by other mental health providers [27,40,41], and hence they are less willing to use them. In addition to private practice setting, CBT theoretical orientation was predictive of willingness to use unguided web-based self-help. This finding may reflect the recent increase

in iCBTs in research and practice settings, many of which are self-guided [42,43]. Increasing knowledge about digital and nondigital LITs and disseminating the interventions that individual providers are willing to use may be a useful way of increasing the reach of LITs [44]. One strategy to increase the use of LITs on treatment waiting lists would be to target the dissemination and implementation of digital and nondigital LITs to individuals who already have positive attitudes toward them—in our study, CBT practitioners and people with private practices. Alternatively, researchers could study interventions to increase the willingness to use digital and nondigital LITs by those not predisposed to using them, for example people who are not CBT practitioners.

## Conclusions

We investigated treatment waiting lists and attitudes toward LITs for patients in waiting lists. Most providers appear to keep a waiting list, but most of them do not provide LITs to individuals on their waiting lists. In general, attitudes toward using LITs for patients in waiting lists were positive. Future research should investigate manipulating attitudes toward digital and nondigital LITs as well as structural barriers that may influence their use. For example, regarding individuals in private practice, who may be less likely to recommend guided LITs, qualitative data from our participants and from other studies [45] highlight questions about providers' legal and ethical liability related to giving LIT guidance for participants on their waiting lists (eg, "I think one might assume a level of risk if participating in guided exercises but not therapy associated with their office"). Additionally, providers may be more, rather than less, likely to recommend LITs to patients with different features (eg, less severe symptoms) [46]. The perceived efficacy of an intervention also seems to be a major correlate of its use, so interventions that educate providers about LIT research are also worth exploring. Future work should clarify the nature of liability when recommending digital and nondigital LITs, as this may be an obstacle to uptake.

---

## Acknowledgments

The authors would like to thank the participants for their time and thoughtful responses to the survey questions.

This research was partially funded by the National Institute of Mental Health (grant T32MH103213; PI: Hetrick), which provided support for AP and grants (KL2TR002530 and UL1TR002529; PI: Shekhar) from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award, which provided support for LLL, as well as the National Institute on Aging (R01AG070931-02; PI: ACK).

---

## Data Availability

No other papers currently use these data. Study data, code, and survey materials can be accessed on the Open Science Framework website [32].

---

## Conflicts of Interest

LLL has received consulting fees from Happify Health Inc, who had no role in this study. The funders had no role in the drafting of the manuscript.

---

Multimedia Appendix 1  
Supplementary Table S1.

[DOCX File , 25 KB - [formative\\_v6i9e39787\\_app1.docx](#) ]

## References

1. Baxter AJ, Vos T, Scott KM, Ferrari AJ, Whiteford HA. The global burden of anxiety disorders in 2010. *Psychol Med* 2014 Jan 22;44(11):2363-2374. [doi: [10.1017/s0033291713003243](#)]
2. James S, Abate D, Abate K, Abay S, Abbafati C, Abbasi N, et al. . [doi: [10.1016/S0140-6736\(18\)32279-7](#)]
3. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet* 2013 Nov;382(9904):1575-1586. [doi: [10.1016/s0140-6736\(13\)61611-6](#)]
4. Greenberg PE, Fournier A, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psychiatry* 2015 Feb 25;76(02):155-162. [doi: [10.4088/jcp.14m09298](#)]
5. Prevention C. Anxiety and depression, household pulse survey. CDC. URL: <https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm> [accessed 2021-12-14]
6. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspect Psychol Sci* 2011 Jan 03;6(1):21-37. [doi: [10.1177/1745691610393527](#)] [Medline: [26162113](#)]
7. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. *Behav Res Ther* 2017 Jan;88:7-18. [doi: [10.1016/j.brat.2016.06.004](#)] [Medline: [28110678](#)]
8. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychol Med* 2013 Aug 09;44(6):1303-1317. [doi: [10.1017/s0033291713001943](#)]
9. Uimonen M, Kuitunen I, Paloneva J, Launonen AP, Ponkilainen V, Mattila VM. The impact of the COVID-19 pandemic on waiting times for elective surgery patients: a multicenter study. *PLoS One* 2021 Jul 6;16(7):e0253875 [FREE Full text] [doi: [10.1371/journal.pone.0253875](#)] [Medline: [34228727](#)]
10. Kutikov A, Weinberg DS, Edelman MJ, Horwitz EM, Uzzo RG, Fisher RI. A war on two fronts: cancer care in the time of COVID-19. *Ann Intern Med* 2020 Jun 02;172(11):756-758. [doi: [10.7326/m20-1133](#)]
11. Brown SA, Parker JD, Godding PR. Administrative, clinical, and ethical issues surrounding the use of waiting lists in the delivery of mental health services. *J Behav Health Serv Res* 2002 May;29(2):217-228. [doi: [10.1007/BF02287708](#)] [Medline: [12032979](#)]
12. Cuijpers P, Karyotaki E, Ciharova M, Miguel C, Noma H, Furukawa TA. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: a meta-analysis. *Acta Psychiatr Scand* 2021 Sep 04;144(3):288-299 [FREE Full text] [doi: [10.1111/acps.13335](#)] [Medline: [34107050](#)]
13. Krendl AC, Lorenzo-Luaces L. Identifying peaks in attrition after clients initiate mental health treatment in a university training clinic. *Psychol Serv* 2022 Aug;19(3):519-526. [doi: [10.1037/ser0000469](#)] [Medline: [34166028](#)]
14. Swift JK, Whipple JL, Sandberg P. A prediction of initial appointment attendance and initial outcome expectations. *Psychotherapy (Chic)* 2012 Dec;49(4):549-556. [doi: [10.1037/a0029441](#)] [Medline: [22889099](#)]
15. Reitzel LR, Stellrecht NE, Gordon KH, Lima EN, Wingate LR, Brown JS, et al. Does time between application and case assignment predict therapy attendance or premature termination in outpatients? *Psychol Serv* 2006 Feb;3(1):51-60. [doi: [10.1037/1541-1559.3.1.51](#)]
16. Steinert C, Stadter K, Stark R, Leichsenring F. The effects of waiting for treatment: a meta-analysis of waitlist control groups in randomized controlled trials for social anxiety disorder. *Clin Psychol Psychother* 2017 May 22;24(3):649-660. [doi: [10.1002/cpp.2032](#)] [Medline: [27445199](#)]
17. Punton G, Dodd AL, McNeill A. 'You're on the waiting list': an interpretive phenomenological analysis of young adults' experiences of waiting lists within mental health services in the UK. *PLoS One* 2022 Mar 18;17(3):e0265542 [FREE Full text] [doi: [10.1371/journal.pone.0265542](#)] [Medline: [35303040](#)]
18. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry* 2017 Mar 26;16(1):90-99 [FREE Full text] [doi: [10.1002/wps.20388](#)] [Medline: [28127925](#)]
19. Cuijpers P. Bibliotherapy in unipolar depression: a meta-analysis. *J Behav Ther Exp Psychiatry* 1997 Jun;28(2):139-147. [doi: [10.1016/s0005-7916\(97\)00005-0](#)]
20. Gregory RJ, Schwer Canning S, Lee TW, Wise JC. Cognitive bibliotherapy for depression: a meta-analysis. *Prof Psychol Res Pr* 2004 Jun;35(3):275-280. [doi: [10.1037/0735-7028.35.3.275](#)]
21. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. *Can J Psychiatry* 2019 Jul 16;64(7):465-470 [FREE Full text] [doi: [10.1177/0706743719839381](#)] [Medline: [31096757](#)]
22. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cogn Behav Ther* 2018 Jan 07;47(1):1-18. [doi: [10.1080/16506073.2017.1401115](#)] [Medline: [29215315](#)]

23. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychol Med* 2010 Apr 21;40(12):1943-1957. [doi: [10.1017/s0033291710000772](https://doi.org/10.1017/s0033291710000772)]
24. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. *JAMA Psychiatry* 2019 Jul 01;76(7):700-707 [FREE Full text] [doi: [10.1001/jamapsychiatry.2019.0268](https://doi.org/10.1001/jamapsychiatry.2019.0268)] [Medline: [30994877](https://pubmed.ncbi.nlm.nih.gov/30994877/)]
25. Williams C. Use of written cognitive-behavioural therapy self-help materials to treat depression. *Adv psychiatr treat* 2018 Jan 02;7(3):233-240. [doi: [10.1192/apt.7.3.233](https://doi.org/10.1192/apt.7.3.233)]
26. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. *JAMA Psychiatry* 2017 May 01;74(4):351-359. [doi: [10.1001/jamapsychiatry.2017.0044](https://doi.org/10.1001/jamapsychiatry.2017.0044)] [Medline: [28241179](https://pubmed.ncbi.nlm.nih.gov/28241179/)]
27. Brantnell A, Woodford J, Baraldi E, van Achterberg T, von Essen L. Views of implementers and nonimplementers of internet-administered cognitive behavioral therapy for depression and anxiety: survey of primary care decision makers in Sweden. *J Med Internet Res* 2020 Aug 12;22(8):e18033 [FREE Full text] [doi: [10.2196/18033](https://doi.org/10.2196/18033)] [Medline: [32784186](https://pubmed.ncbi.nlm.nih.gov/32784186/)]
28. Schleider JL, Sung J, Bianco A, Gonzalez A, Vivian D, Mullarkey M. Open pilot trial of a single-session consultation service for clients on psychotherapy wait-lists. *PsyArXiv Preprint* published online in Jan 15, 2021. [doi: [10.31234/osf.io/fdwqk](https://doi.org/10.31234/osf.io/fdwqk)]
29. Levin ME, Hicks ET, Krafft J. *J Am Coll Health* 2022 Jan 09;70(1):165-173. [doi: [10.1080/07448481.2020.1728281](https://doi.org/10.1080/07448481.2020.1728281)] [Medline: [32150519](https://pubmed.ncbi.nlm.nih.gov/32150519/)]
30. Hentati A, Forsell E, Ljótsson B, Lindfors N, Kraepelien M. A self-guided and monitored digital problem-solving intervention for patients with symptoms of depression or anxiety on the waiting list for treatment in routine psychiatric care: feasibility study. *BJPsych Open* 2022 Feb 08;8(2):e43 [FREE Full text] [doi: [10.1192/bjo.2022.14](https://doi.org/10.1192/bjo.2022.14)] [Medline: [35130998](https://pubmed.ncbi.nlm.nih.gov/35130998/)]
31. Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process* 1991 Dec;50(2):179-211. [doi: [10.1016/0749-5978\(91\)90020-t](https://doi.org/10.1016/0749-5978(91)90020-t)]
32. Peipert A. Waiting lists for psychotherapy and provider attitudes towards low-intensity interventions. *PsyArXiv Preprint* posted online May 3, 2022. [doi: [10.31234/osf.io/63gfs](https://doi.org/10.31234/osf.io/63gfs)]
33. Tang F, Ishwaran H. Random forest missing data algorithms. *Stat Anal Data Min* 2017 Dec 13;10(6):363-377 [FREE Full text] [doi: [10.1002/sam.11348](https://doi.org/10.1002/sam.11348)] [Medline: [29403567](https://pubmed.ncbi.nlm.nih.gov/29403567/)]
34. The R language and environment for statistical computing.: R Core Team; 2017. URL: <https://www.R-project.org> [accessed 2022-05-22]
35. Lin L, Stamm K, Christidis P. Demographics of the US psychology workforce: findings from the 2007-16 American Community Survey. *APA* 2018:3-12. [doi: [10.1037/e506742018-001](https://doi.org/10.1037/e506742018-001)]
36. Norcross JC, Karpik CP. Clinical psychologists in the 2010s: 50 years of the APA division of clinical psychology. *Clin Psychol* 2012 Mar;19(1):1-12. [doi: [10.1111/j.1468-2850.2012.01269.x](https://doi.org/10.1111/j.1468-2850.2012.01269.x)]
37. Norcross JC, Sayette MA, Pomerantz AM. Doctoral training in clinical psychology across 23 years: Continuity and change. *J Clin Psychol* 2018 Mar 20;74(3):385-397. [doi: [10.1002/jclp.22517](https://doi.org/10.1002/jclp.22517)] [Medline: [29156100](https://pubmed.ncbi.nlm.nih.gov/29156100/)]
38. Steinman KJ, Shoben AB, Dembe AE, Kelleher KJ. How long do adolescents wait for psychiatry appointments? *Community Ment Health J* 2015 Oct 25;51(7):782-789. [doi: [10.1007/s10597-015-9897-x](https://doi.org/10.1007/s10597-015-9897-x)] [Medline: [26108305](https://pubmed.ncbi.nlm.nih.gov/26108305/)]
39. Blech B, West JC, Yang Z, Barber KD, Wang P, Coyle C. Availability of network psychiatrists among the largest health insurance carriers in Washington, D.C. *Psychiatr Serv* 2017 Oct 01;68(9):962-965. [doi: [10.1176/appi.ps.201600454](https://doi.org/10.1176/appi.ps.201600454)] [Medline: [28457210](https://pubmed.ncbi.nlm.nih.gov/28457210/)]
40. Miller KE, Kuhn E, Yu J, Owen JE, Jaworski BK, Taylor K, et al. Use and perceptions of mobile apps for patients among VA primary care mental and behavioral health providers. *Prof Psychol Res Pr* 2019 Jun;50(3):204-209. [doi: [10.1037/pro0000229](https://doi.org/10.1037/pro0000229)]
41. Cliffe B, Croker A, Denne M, Stallard P. Clinicians' use of and attitudes towards technology to provide and support interventions in child and adolescent mental health services. *Child Adolesc Ment Health* 2020 May 11;25(2):95-101. [doi: [10.1111/camh.12362](https://doi.org/10.1111/camh.12362)] [Medline: [32307835](https://pubmed.ncbi.nlm.nih.gov/32307835/)]
42. Aguilera A, Muench F. There's an app for that: information technology applications for cognitive behavioral practitioners. *Behav Ther (N Y N Y)* 2012 May;35(4):65-73 [FREE Full text] [Medline: [25530659](https://pubmed.ncbi.nlm.nih.gov/25530659/)]
43. Wasil AR, Palermo EH, Lorenzo-Luaces L, DeRubeis RJ. Is there an app for that? A review of popular apps for depression, anxiety, and well-being. *Cogn Behav Pract* 2021 Oct (forthcoming). [doi: [10.1016/j.cbpra.2021.07.001](https://doi.org/10.1016/j.cbpra.2021.07.001)]
44. Michie S, Atkins L, West R. *The Behaviour Change Wheel: A guide to Designing Interventions*. Great Britain: Silverback Publishing; 2014.
45. Borgueta AM, Purvis CK, Newman MG. Navigating the ethics of Internet-guided self-help interventions. *Clin Psychol (New York)* 2018 Jul;25(2):e12235 [FREE Full text] [doi: [10.1111/cpsp.12235](https://doi.org/10.1111/cpsp.12235)] [Medline: [33664555](https://pubmed.ncbi.nlm.nih.gov/33664555/)]
46. Lorenzo-Luaces L, DeRubeis RJ, Bennett IM. Primary care physicians' selection of low-intensity treatments for patients with depression. *Fam Med* 2015;47(7):511-516. [Medline: [26562637](https://pubmed.ncbi.nlm.nih.gov/26562637/)]



## Abbreviations

**CBT:** cognitive behavioral therapy  
**CMD:** common mental disorder  
**DMHI:** digital mental health intervention  
**iCBT:** internet-based cognitive behavioral therapy  
**LIT:** low-intensity treatment  
**OR:** odds ratio

*Edited by A Mavragani; submitted 23.05.22; peer-reviewed by S Schueller, K Matsumoto; comments to author 03.07.22; revised version received 19.07.22; accepted 22.08.22; published 16.09.22.*

*Please cite as:*

*Peipert A, Krendl AC, Lorenzo-Luaces L*

*Waiting Lists for Psychotherapy and Provider Attitudes Toward Low-Intensity Treatments as Potential Interventions: Survey Study*

*JMIR Form Res 2022;6(9):e39787*

*URL: <https://formative.jmir.org/2022/9/e39787>*

*doi: [10.2196/39787](https://doi.org/10.2196/39787)*

*PMID: [36112400](https://pubmed.ncbi.nlm.nih.gov/36112400/)*

©Allison Peipert, Anne C Krendl, Lorenzo Lorenzo-Luaces. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 16.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Influence of 2 Digital Exercise Modules of a Multimodular System on Balance and Leg Strength Under Consideration of Use Adherence: Prospective Cohort Study

Verena Venek<sup>1\*</sup>, DI, BSc; Christina Kranzinger<sup>1\*</sup>, Mag, MSc, BStat; Sonja Jungreitmayr<sup>2</sup>, Mag; Susanne Ring-Dimitriou<sup>2</sup>, Prof Dr Dr; Hermann Schwameder<sup>2</sup>, Prof Dr; Thomas Stöggel<sup>2,3</sup>, Prof Dr

<sup>1</sup>Salzburg Research Forschungsgesellschaft mbH, Salzburg, Austria

<sup>2</sup>Department of Sport and Exercise Science, Paris Lodron University Salzburg, Salzburg, Austria

<sup>3</sup>Red Bull Athletes Performance Center, Salzburg, Austria

\*these authors contributed equally

**Corresponding Author:**

Verena Venek, DI, BSc

Salzburg Research Forschungsgesellschaft mbH

Jakob Haringer Straße 5/3

Salzburg, 5020

Austria

Phone: 43 6622288 ext 415

Email: [v.venek@fh-kaernten.at](mailto:v.venek@fh-kaernten.at)

## Abstract

**Background:** To empower healthy aging, digital solutions embed multiple modules for physical activity, cognitive health promotion, and social engagement. Integrating new empowering technologies such as digital exercise monitoring requires assessment measures and analysis procedures, considering variable compliance of users with different modules.

**Objective:** This study aims to assess the influence of a tablet-based and a feedback system-based exercise module on balance and leg strength by considering use adherence instead of the use of the entire multimodular system.

**Methods:** In the prospective cohort study within the fit4AAL project, 83 users (n=67, 81% women; n=16, 19% men; mean age 66.2, SD 2.3 years) used the 2 digital exercise modules of a multimodular physical activity promotion system for >18 weeks. A data-driven clustering method based on the average use frequency of the exercise modules determined the number of user types that met the World Health Organization-recommended training frequency of at least twice per week. On the basis of this use adherence, statistical analysis was performed with features of functional performance tests (unipedal stance, 30-second chair rise, Y-balance, and hurdle step tests). The tests were conducted 6 months before the intervention, immediately before the intervention, and after the intervention, comparing the baseline phase with the 3 feedback use groups of the study (using only the tablet, the tablet and the feedback system, or only the feedback system).

**Results:** Of the 83 users, 43 (52%) met the World Health Organization-recommended frequency of muscle-strengthening activities. Overall, the feedback use groups achieved, on average, more chair rises in 30 seconds than the baseline group ( $P=.01$ ; moderate effect size of 0.07). Of the 43 users, 26 (60%) additionally used the feedback system-based exercise module. They improved in balance compared with the users using either the tablet or the feedback system ( $P=.02$ ). In addition, they improved their leg strength within the group ( $P=.04$ ) and compared with the baseline ( $P=.01$ ).

**Conclusions:** The additional use of a feedback system showed a tendency to positively maintain and influence the already exceptionally high functional performance of older adults. Considering use adherence in future multimodular system studies is crucial to assess the influence of single and combined use of exercise modules on functional performance.

(*JMIR Form Res* 2022;6(9):e36805) doi:[10.2196/36805](https://doi.org/10.2196/36805)

**KEYWORDS**

active and assisted living; functional fitness training; information and communication technology; use adherence

## Introduction

The things we do and value in life affect how we age. The ultimate goal is to age healthily and, therefore, increase the number of healthy years of life. The World Health Organization (WHO) describes healthy aging as “the process of developing and maintaining the functional ability that enables wellbeing in older age” [1]. Functional ability encompasses the intrinsic capacity of people’s mental and physical abilities to cope with daily life. Physical abilities affect mental abilities and vice versa [2]. In the field of health promotion, physical activity has been prescribed as a preventive measure for reducing risks of functional ability and noncommunicable diseases [3].

To engage older people and promote physical activity, new empowering technologies have been considered to support compliance with training at home. In comparison with assistive technologies, which aim to alleviate the effects of disabilities such as immobility, empowering technologies aim to help people prevent functional disability [4]. For the users, it has to be an experience to maintain or even increase compliance with exercise and make the long-term effects tangible. Therefore, digital exercise modules have included different technologies and devices, such as fitness trackers and camera systems, to provide feedback on exercise performance and movement quality during training sessions.

Although interventions with exercise feedback and monitoring have been used with positive results on functional performance, procedures to thoroughly report and test the digital interventions and their adherence for older adults are lacking [5]. This assessment is even more challenging when, in particular, older participants train unsupervised [6]. Nevertheless, this setting is common in the field of Active and Assisted Living (AAL). For example, AAL projects funded by the AAL Joint Programme of the European Union should integrate a proof of concept, market potential, and strategies of the solutions, as well as impact evaluations, into 3- to 4-year AAL projects [7]. Hence, studies evaluating the effects of technology-assisted physical activity interventions in real-life settings are required [8,9].

The goal of AAL research is to promote active and healthy aging by developing and investigating age-appropriate services and solutions based on Information and Communication Technology (ICT). The developed solutions are often ICT-based multimodal systems that address  $\geq 1$  component of successful aging; that is, chronic disease management, maintenance of physical and cognitive health, and active social engagement [10]. For physical health, the WHO recommends regular muscle-strengthening training involving major muscle groups for the age group of 18 to 64 years on  $\geq 2$  days a week [11]. To achieve the recommended training frequency, digital exercise modules integrated into multimodal solutions have been considered to augment usual coaching, reduce costs, and increase training accessibility to support people in establishing training routines [12-14].

Despite proven long-term effects of regular exercise on functional ability, the impact measures and analysis procedures for evidence of the usefulness of digital exercise modules vary. Although the use adherence in the ZentrAAL project had been

analyzed independently from functional performance [15], CareInMovement introduced an expert-based clustering of user types to identify use-related differences [16]. However, the workouts provided in both projects were accessible via a tablet app, and the effects on participants’ functional performance related to the fitness program were investigated rather than the impact of the technological devices themselves on physical capabilities. Dasgupta et al [10] surveyed tablet-based intervention studies, which were characterized by limited sample sizes and durations, and revealed significant improvements in functional performance in terms of gait velocity, repeated chair rises, and static balance. However, the mentioned apps mainly focused on providing training content without feedback or monitoring exercise execution.

A study protocol given by Belleville et al [17] described the intended outcome analysis for the tablet-based multimodal solution *StayFitLonger* related to physical and cognitive health training at home. Performance on the “Timed-Up & Go” test was selected as the primary outcome measure of physical health, which is assessed before and after the 26-week intervention. Use adherence was mentioned to be considered in future evaluations to determine whether the recommended dose, volume, and frequency of the physical activity program could be maintained over time and to evaluate the efficacy.

A similar multimodal AAL solution for physical activity promotion was developed in the Austrian AAL project, fit4AAL. The purpose of the modules was not only to support daily physical activity and regular exercising but also to promote knowledge about an active and assisted lifestyle. The fit4AAL field trial included 2 phases using a randomized controlled trial design with a waitlist baseline group [18]. In the first trial phase with a randomized-assigned intervention group and a baseline group, the impact of the entire solution was investigated. For example, the exercise modules improved muscular strength and flexibility in older women in the first phase [19]. In the second trial phase, the same but matured digital intervention was applied to the baseline group. In both phases, the AAL solution integrated the 2 digital exercise modules. A tablet-based version and a feedback system-based version of a personalized training program focused on improving functional abilities and establishing a training routine.

To evaluate the influence of the digital exercise modules of the multimodal solution on balance and leg strength, including the movement quality of older adults, we hypothesized that the additional use of the feedback system-based exercise module would improve the functional performance of the participants when they met the WHO-recommended training frequency. As the feedback system-based exercise module was more mature in the second phase of the field trial, this prospective cohort study investigated the data from the participants of the waitlist baseline group of the fit4AAL field trial. Moreover, one of the questions was how many participants met the WHO-recommended weekly training frequency during the intervention phase. Therefore, use adherence was used to investigate the compliance with the 2 exercise modules.

## Methods

### Digital Exercise Modules and Training Program

The multimodal AAL solution comprised 4 modules: an e-learning module, an activity-tracking module, and 2 exercise modules forming the digital home training. One of the two personalized digital exercise modules was accessible via a tablet-based Android app version, including the other modules as well. The other exercise module provided live feedback for older adults during training via the feedback system-based Android app version. This feedback system monitored exercise through skeleton tracking of the 3D camera system Orbbec Persee (Orbbec) connected to the users' television monitors. This version, as well as the tablet-based exercise module, offers training sessions tailored to users' fitness levels, with 365 functional exercises [20].

The used multidimensional training programs focused on varied but structured movements to maintain or even increase functional fitness [21]. All training sessions comprised a warm-up, main, and cool-down phase. Exercises to mobilize and invigorate the cardiovascular system, such as shoulder circles and marching in place, started the warm-up. The main part included strengthening exercises such as squats and table push-ups. Stretching exercises for the lower and upper body concluded each training session. The users could select among 10-minute, 20-minute, or 30-minute training sessions, and the suggested workouts changed daily. At the start of each exercise, a thorough description of the exercise in written, video-based, and spoken form was available.

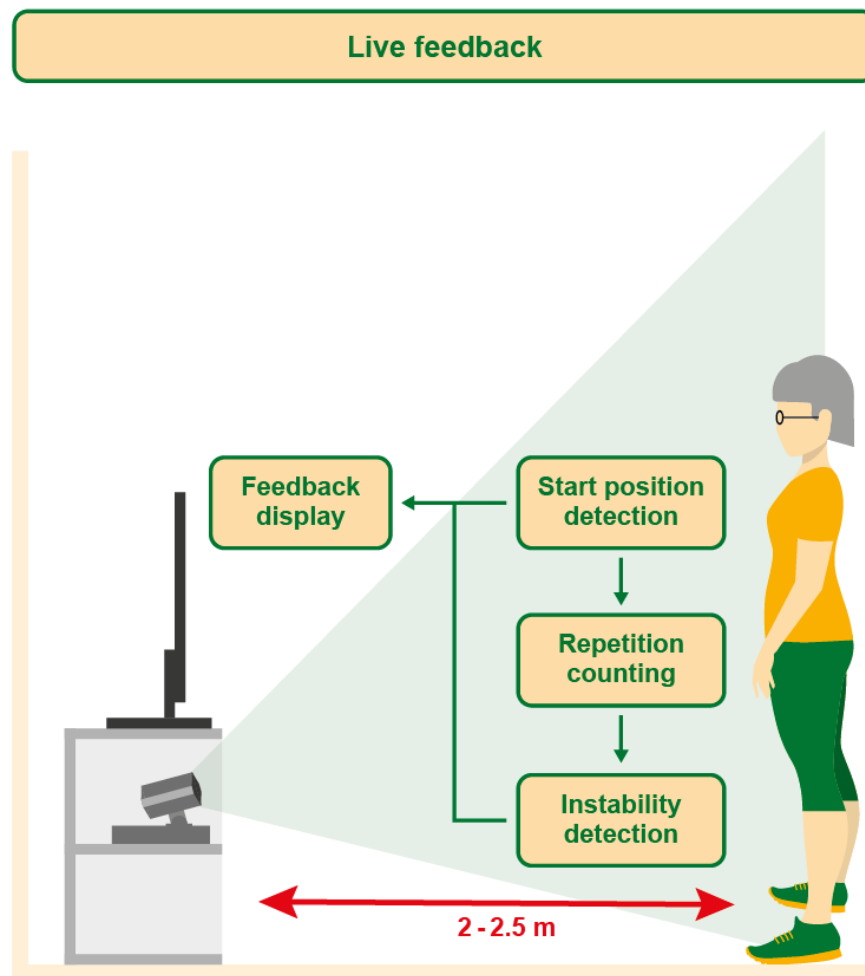
The modules differed after the exercise description. In the exercise module on the tablet, the exercise description video and a countdown were displayed while the users were performing the exercise for the prescribed amount of time. For each exercise, users had the possibility to either skip the exercise

or confirm it by clicking on the corresponding buttons in the app. In both cases, the next exercise was described. After the last exercise, an overview of the workout, including the actual workout duration and number of performed exercises, was displayed.

Additional live feedback during exercise execution was available in the feedback system-based exercise module. To use skeleton tracking (Nuitrack SDK, version 1.3.1; 3DiVi Inc), the users were advised to stand 2 to 2.5 meters away and in front of the Persee (Figure 1). In addition to the exercise description video, a camera live stream mirrored users so that they could visually adjust and correct their posture. Furthermore, several exercises were tagged with feedback-providing algorithms: from the 365 exercises, 154 were tagged with one of the start position detection algorithms, 122 were tagged with one of the repetition counting algorithms, and 100 were tagged with the instability detection algorithms.

Repetition counting and instability detection required start position detection. Start position detection algorithms supported users in taking the required pose for an exercise. For example, squats were counted only when users took the required starting position of hip-wide standing. Repetition counting algorithms enabled the users to follow the number of repetitions they achieved, such as the number of squats they performed. Additional instability detection algorithms monitored the correct postures during exercise execution. For example, during the one-leg stand exercises, the users were visually notified when their sway to either side left a balanced stance. As the skeleton tracking of the 3D camera system was used, the focus was on exercises starting from the standing or sitting position [22]. This excluded exercises in prone or supine positions performed on the floor or leaning against a wall, as well as exercises where users faced the camera laterally. The camera live stream in the app was available for all exercises.

**Figure 1.** Live feedback- and exercise-monitoring functionalities of the feedback system–based exercise module and recommended distance from the 3D camera system of 2 to 2.5 meters.



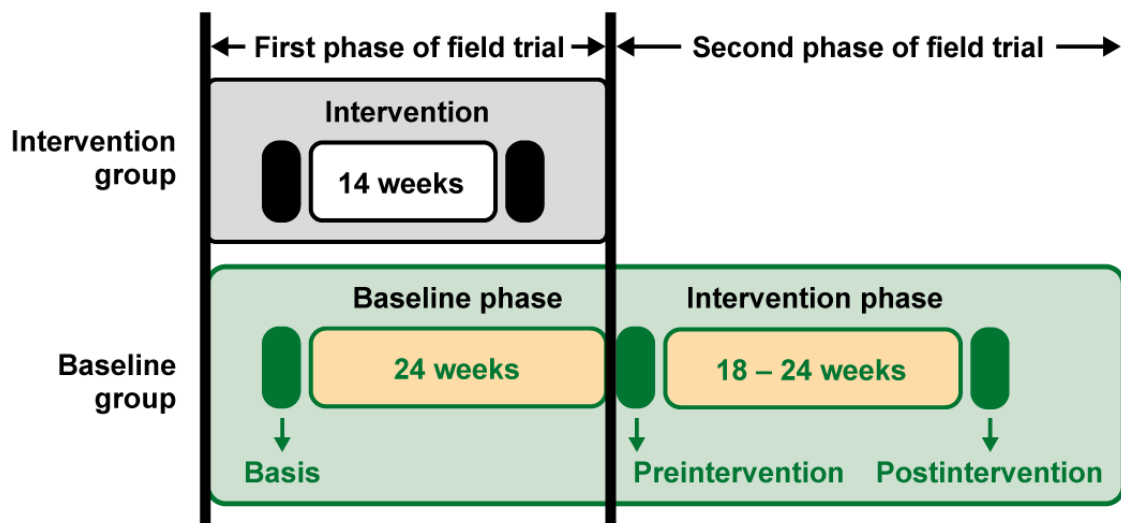
## Participants

The thorough recruitment process of fit4AAL participants in Austria was described by Trukeschitz et al [18]. It was separated into three stages: potential participants (1) received letters of invitation via mail, magazines, newspapers, websites, and digital newsletters; (2) had to confirm their interest following a developed questionnaire on the web or by telephone; and (3) were selected by the project team based on the questionnaire answers. The questionnaire was used to identify individuals who had been retired for 2.5 to 6 years and were willing to participate in a scientific study of physical activity promotion, were not dependent on mobility aids, had no chronic diseases or physical limitations, did not have a personal trainer at the time of recruitment, or exercised  $>4$  times per week. Furthermore, participants who were technology savvy were selected based on possession of an email address, a monitor, and free space (approximately 2-2.5 meters) in front of the monitor to exercise.

The selected study participants were randomly assigned to the intervention and baseline groups of the first fit4AAL trial phase. The baseline group of the first trial phase became the intervention group of the second trial phase because of the randomized controlled trial with a waitlist baseline group design [15]. This means that the multimodal AAL solution was first applied to the intervention group and then to the baseline group. For the prospective cohort study of the baseline group, the baseline and intervention phases of the baseline group were investigated (Figure 2).

Excluding dropouts and nonusers from the 109 participating adults, 91 (83%) used at least one of the four modules of the multimodal AAL solution during the intervention phase [23]. Of the 91 users, 86 (94%) used either 1 of the 2 or both digital exercise modules. This resulted in an adherence rate of 94.5% to the exercise modules of the AAL solution. From these 86 users, 3 (3%) were removed as they did not show up for the tests or they had not finished any workout. The average age of the 83 users ( $n=67$ , 81% women and  $n=16$ , 19% men) was 66.2 (SD 2.3) years.

**Figure 2.** Field trial phases with a randomly assigned intervention group and waitlist baseline group. The trial phases of the baseline group were separated into baseline and intervention phases, including 3 functional performance assessments (basis, preintervention, and postintervention).



### Ethics Approval

All study participants were informed about their rights, data use, and exit strategies by signing an informed consent form before the study. The study design was positively evaluated by the ethics committee of the University of Salzburg (EK-GZ: 09/2018). The videos of the feedback system-based module were neither saved nor sent to any servers to ensure the participants' privacy. The skeleton tracking was performed on the depth images exclusively on the system and did not require any video data.

### Study Design

A prospective cohort study was conducted within a year. Participants performed functional performance tests supervised by sports scientists for basis and pre- and postintervention measurements. The functional performance tests included the unipedal stance (UPS) test, the 30-second chair rise test (CRT), the Y-balance test (YBT), and the hurdle step (HS) test.

Furthermore, 6 months before the intervention, the basis measurements were collected within 4 to 8 weeks because of staged appointment coordination. The data from the basis and preintervention assessment points defined the baseline phase. The preintervention functional tests assessed the fitness level of the participants for the configuration of the training program. For the following 18 to 24 weeks, the participants received the multimodal AAL solution with the digital exercise modules for their homes. After the intervention, the participants returned to the sports scientists for the postintervention measurements. [Figure 2](#) shows the timeline of the study.

During the intervention phase, the workout information and app use were monitored. The workout data comprised the workout duration in hours, the workout duration performed with the feedback system, and the number of completed workouts. Furthermore, the workout data included the number of times the 10-minute, 20-minute, or 30-minute workouts were selected.

App use data were recorded by digitally logging the visits of the modules of the AAL solution via the open-source web

analytics tool (Matomo, InnoCraft). A visit was defined as continuous viewing of either one of the exercise modules without breaks for >30 minutes. Aggregated use data included information on visits per day and whether users viewed either the tablet-based or the feedback system-based exercise module.

### Descriptive Analysis of Use Adherence and Feedback Use Groups

Voluntary use at home was expected to influence the compliance with the modules, as well as with the exercise modules. Thus, use adherence was defined in visits to either one of the two exercise modules, determining how often the participants used the exercise modules on average during the intervention. To determine the user types, the Jenks natural breaks cluster algorithm was applied based on the average frequency of visits to the exercise modules to define the interval limits, indicating the use adherence of different user types to the exercise modules in visits per day. The limits multiplied by 7 days estimated the number of visits per week.

The statistical analysis required the distribution of users to the different versions of the exercise modules, resulting in 3 feedback use groups: they described the number of users who trained only with the tablet-based exercise module, those who trained additionally with the feedback system-based exercise module, and those who trained only with the feedback system. For each of the 3 feedback use groups, descriptive statistics, including age, sex, average workout data, and baseline test results, were determined. The reported sex categories were male and female. In addition, the user types with upper interval limits smaller than 2 visits per week were rejected as they did not meet the WHO-recommended amount of muscle-strengthening activities of at least twice per week. Thus, descriptive statistics and statistical analysis were repeated for the adapted feedback use groups, considering use adherence.

### Outcome Measures

#### Balance

The UPS test represents and assesses static balance by recording the time in seconds the participant is able to stand on one leg

without relying on the standing leg [24,25]. The maximal duration of the UPS was set to 60 seconds and the times were noted in seconds to the nearest tenth. Of the 3 attempts for each leg, the maximal durations for both legs (UPS maximum) and each leg were determined (UPS left leg maximum and UPS right leg maximum).

### Leg Strength

The CRT represents and assesses the lower body strength by counting the number of chair rises that can be completed within 30 seconds [26]. A correct chair rise starts with the participant sitting on the chair with arms crossed at chest height, shoulder wide stance, and feet in full contact with the ground, and it ends in a standing position with the hips and knees fully extended.

### Balance and Leg Strength

The YBT assesses the stability, range of motion, strength coordination, and mobility of the upper and lower body for both the left and right sides [27]. It represents dynamic balance and leg strength. The composite reach distances or composite scores for the left leg (YBT left leg) and right leg (YBT right leg) are the sums of the 3 reach directions (anterior, posteromedial, and posterolateral) divided by 3 times the limb length. Participants had to maintain balance on their dominant leg while they were encouraged to reach indicators in the 3 directions as far as possible. Furthermore, a leg symmetry index (YBT leg symmetry) was determined by calculating the difference between the composite reach distance of the left and right leg.

### Movement Quality

The HS is 1 of the 7 fundamental movement patterns of the Functional Movement Screen tests and assesses the functional symmetry by scoring the performed HS representing a measure of human movement quality related to balance and leg strength [28]. The categorical scores were 1, 2, or 3, with a score of 3 expressing the best performance and a score of 1 expressing the worst. The HS test was attempted 3 times, with the highest scores used for the analysis for each leg (HS left leg and HS right leg).

### Statistical Analysis

Statistical analyses were performed using the statistical software R (version 4.0.3; R Foundation for Statistical Computing) [29] and the *rstatix* (version 0.6.0) package [30]. The significance level was set to  $\alpha < .05$ . To determine significance within the groups, a paired Wilcoxon signed-rank sum test was used on the pre- and postintervention data of the feedback use groups and on the basis and preintervention data from the baseline group.

To assess the influence of the training with the digital exercise modules on balance and leg strength, we used the difference between pre- and postintervention measurements from the feedback use groups and the difference between basis and preintervention measurements as control. The Shapiro-Wilk

test was used to screen for normality. As the data set was not normally distributed, the nonparametric Kruskal-Wallis rank-sum test was used to determine the interaction effects between the feedback use groups and the baseline values. Where significant differences among the groups could be determined, post hoc pairwise comparisons using the Wilcoxon signed-rank sum test were performed.

The resulting *P* values were adjusted by using the multiple testing correction method of Benjamini and Hochberg [31]. The  $\eta^2$  value based on the H-statistic was used as a measure of the effect size. A small effect is indicated by values  $< 0.06$ , a moderate effect is indicated by values between 0.06 and 0.14, and a large effect is indicated by values  $\geq 0.14$  [30].

## Results

### Use Adherence and Feedback Use Groups

The descriptive statistics of the feedback use groups are shown in Table 1 as mean values and SD unless otherwise stated.

Of the 83 users, 37 (45%) users exercised with the tablet-based exercise module without live feedback, and 36 (43%) users trained additionally with the feedback system. The remaining 12% (10/83) of participants used only the live feedback system to exercise. Of the 83 users, 4 user type clusters were determined: 40 (48%) users with up to 1.89 visits per week, 21 (25%) users with 1.96 to 4.41 visits per week, 19 (23%) users with 4.48 to 7.42 visits per week, and 3 (4%) users with 7.49 to 14.56 visits per week. Their distribution to the feedback use groups is shown in the Sankey diagram in Figure 3.

The use adherence of the first cluster with 48% (40/83) of the users did not meet the required 2 visits per week upper limit. Three-quarters of the users in this user type used only 1 of the 2 exercise modules, whereas 25% (10/40) used the tablet and the feedback system. The feedback use groups were adapted by rejecting this user type with 40 users, resulting in a total of 43 users. The descriptive statistics of the adapted feedback use groups are presented in Table 2. Looking at the distribution of the feedback use groups considering the use adherence, 40% (17/43) of the remaining users used either one of the exercise modules, and 60% (26/43) used the feedback system-based exercise module in addition to the tablet-based exercise module.

Investigating the workout information, on average 51% (868/1702) and 35% (596/1702) of the workouts in the feedback use group using the tablet and the feedback system were selected as 20-minute and 30-minute workouts, respectively. In comparison, users who exercised with only the tablet-based exercise module chose the 10-minute workout more often on average. In total, 1064 workouts were completed in the feedback use group using only the tablet, 1702 in the feedback use group using the tablet and the feedback system, and 242 workouts were completed in the feedback use group using only the feedback system.

**Table 1.** Descriptive characteristics of the baseline and feedback use groups without considering use adherence, including functional performance test results from baseline data and preintervention data for the feedback use groups (N=83).

Feedback use group	Baseline		Using only the tablet		Using the tablet and the feedback system		Using only the feedback system	
	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)
Age (years)	83 (100)	66.2 (2.3)	37 (45)	65.9 (2.3)	36 (43)	66.3 (2.2)	10 (12)	66.8 (2.8)
Female	67 (81)	N/A <sup>a</sup>	30 (36)	N/A	30 (36)	N/A	7 (8)	N/A
Male	16 (19)	N/A	7 (8)	N/A	6 (7)	N/A	3 (4)	N/A
Workout duration (hours)	— <sup>b</sup>	—	37 (45)	12.7 (15.7)	36 (43)	24.9 (18.5)	10 (12)	11.2 (13.7)
Workout duration with feedback (hours)	—	—	NR <sup>c</sup>	NR	36 (43)	14.1 (16.7)	10 (12)	11.2 (13.7)
Number of completed workouts	—	—	37 (45)	28.8 (32.2)	36 (43)	47.3 (33.7)	10 (12)	24.2 (29.9)
10-minute workouts (%)	—	—	37 (45)	44 (36)	36 (43)	14 (14)	10 (12)	51 (46)
20-minute workouts (%)	—	—	37 (45)	32 (31)	36 (43)	51 (28)	10 (12)	42 (42)
30-minute workouts (%)	—	—	37 (45)	24 (32)	36 (43)	35 (29)	10 (12)	6 (12)
UPS <sup>d</sup> maximum (seconds)	83 (100)	50.8 (16.2)	37 (45)	49.7 (17.7)	36 (43)	56.6 (9.0)	9 (11)	51.4 (15.9)
UPS left leg maximum (seconds)	82 (99)	45.1 (19.9)	37 (45)	47.4 (19.5)	36 (43)	51.7 (16.5)	9 (11)	47.3 (19.1)
UPS right leg maximum (seconds)	83 (100)	47.9 (18.3)	37 (45)	45.6 (18.9)	36 (43)	56.6 (9.0)	9 (11)	50.2 (15.9)
30-second CRT <sup>e</sup> (chair rises)	80 (96)	16.0 (4.7)	37 (45)	15.3 (3.5)	36 (43)	17.6 (4.5)	9 (11)	15.6 (3.0)
YBT <sup>f</sup> left leg composite score	71 (86)	81.7 (11.2)	36 (43)	79.6 (13.4)	36 (43)	85.9 (12.1)	8 (10)	81.2 (10.9)
YBT right leg composite score	71 (86)	80.4 (13.2)	36 (43)	78.7 (12.8)	36 (43)	85.2 (8.7)	8 (10)	82.2 (10.7)
YBT leg symmetry	71 (86)	3.9 (6.2)	36 (43)	3.5 (2.4)	36 (43)	3.4 (4.6)	8 (10)	2.2 (2.4)
HS <sup>g</sup> left leg	79 (95)	2.2 (0.6)	37 (45)	2.1 (0.7)	36 (43)	2.3 (0.6)	8 (10)	2.1 (0.4)
HS right leg	79 (95)	2.1 (0.7)	37 (45)	2.1 (0.7)	36 (43)	2.3 (0.7)	8 (10)	2.3 (0.5)

<sup>a</sup>N/A: not applicable (no mean or SD values for sex categories applicable).

<sup>b</sup>Not available since baseline group did not receive any of the exercise modules.

<sup>c</sup>NR: not reported (not available since feedback use group using only the tablet did not perform exercises with the feedback system).

<sup>d</sup>UPS: unipedal stance.

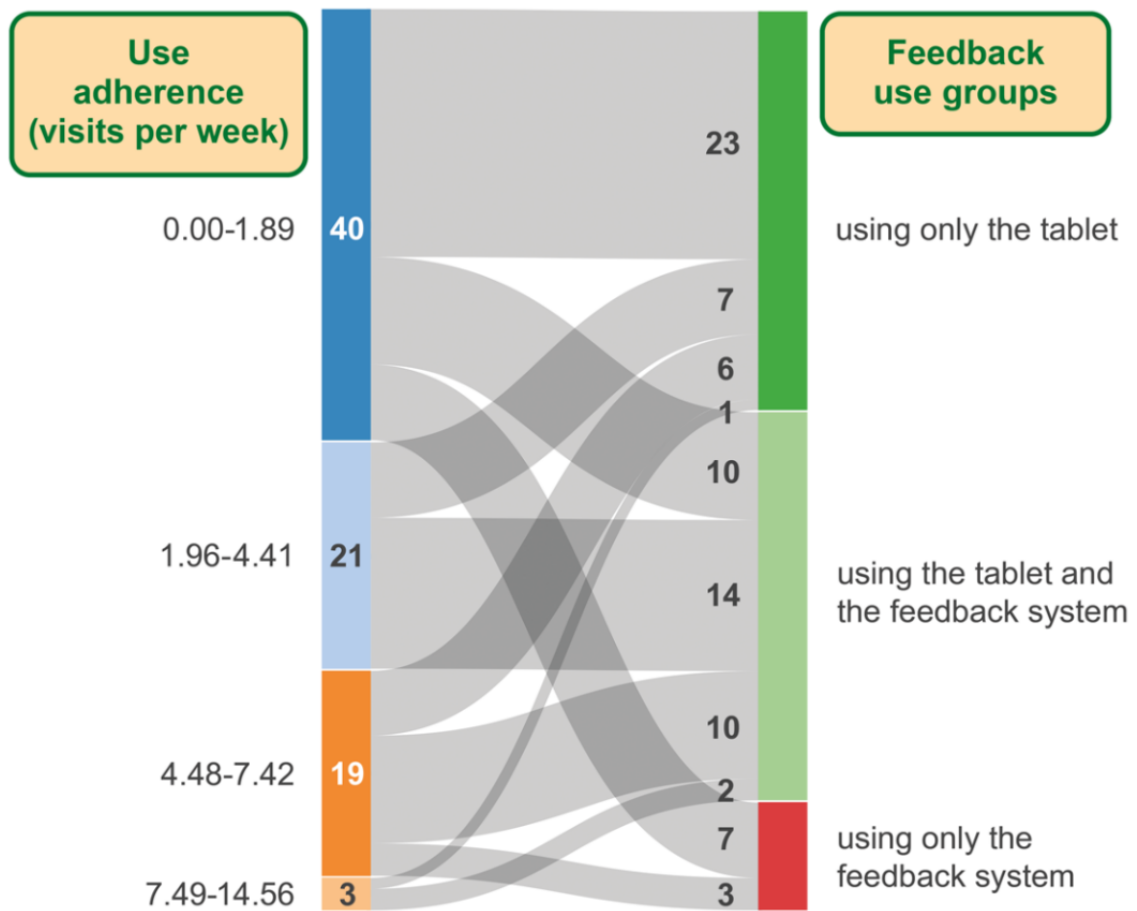
<sup>e</sup>CRT: chair rise test.

<sup>f</sup>YBT: Y-balance test.

<sup>g</sup>HS: hurdle step.



**Figure 3.** Sankey diagram showing the user flows of user types to the feedback use groups (only tablet-based exercise module used, tablet-based and feedback system–based exercise modules used, and only feedback system–based exercise module used) of the investigated 83 study participants.



**Table 2.** Descriptive characteristics of the feedback use groups considering use adherence (from 0.28 visits per day), including functional performance test results from preintervention data for the feedback use groups (N=43).

Feedback use group	Using only the tablet		Using the tablet and the feedback system		Using only the feedback system	
	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)
Age (years)	14 (33)	64.9 (1.0)	26 (60)	66.0 (2.3)	3 (7)	66.0 (1.0)
Female	13 (30)	N/A <sup>a</sup>	22 (51)	N/A	3 (7)	N/A
Male	1 (2)	N/A	4 (9)	N/A	0 (0)	N/A
Workout duration (hours)	14 (33)	27.1 (17.1)	26 (60)	31.5 (17.7)	3 (7)	28.7 (6.7)
Workout duration with feedback (hours)	— <sup>b</sup>	—	26 (60)	17.5 (18.5)	3 (7)	28.7 (6.7)
Number of completed workouts	14 (33)	59.6 (32.9)	26 (60)	60.0 (31.2)	3 (7)	63.0 (20.4)
10-minute workouts (%)	14 (33)	34 (32)	26 (60)	15 (16)	3 (7)	4 (2)
20-minute workouts (%)	14 (33)	33 (28)	26 (60)	51 (28)	3 (7)	95 (2)
30-minute workouts (%)	14 (33)	33 (35)	26 (60)	35 (29)	3 (7)	1 (1)
UPS <sup>c</sup> maximum (seconds)	14 (33)	59.0 (3.7)	26 (60)	58.0 (6.4)	3 (7)	60.0 (0.0)
UPS left leg maximum (seconds)	14 (33)	57.7 (5.9)	26 (60)	53.7 (15.2)	3 (7)	60.0 (0.0)
UPS right leg maximum (seconds)	14 (33)	51.9 (14.1)	26 (60)	57.8 (6.5)	3 (7)	60.0 (0.0)
30-second CRT <sup>d</sup> (chair rises)	14 (33)	15.5 (3.3)	26 (60)	18.2 (4.6)	3 (7)	16.3 (3.3)
YBT <sup>e</sup> left leg composite score	14 (33)	85.2 (6.1)	26 (60)	85.3 (13.4)	3 (7)	81.6 (8.2)
YBT right leg composite score	14 (33)	83.2 (7.6)	26 (60)	84.7 (9.0)	3 (7)	83.7 (5.8)
YBT leg symmetry	14 (33)	3.3 (2.6)	26 (60)	3.7 (5.1)	3 (7)	2.1 (2.4)
HS <sup>f</sup> left leg	14 (33)	2.0 (0.7)	26 (60)	2.3 (0.6)	3 (7)	2.0 (0.0)
HS right leg	14 (33)	2.1 (0.5)	26 (60)	2.2 (30.7)	3 (7)	2.3 (0.6)

<sup>a</sup>N/A: not applicable (no mean or SD values for sex categories applicable).

<sup>b</sup>Not available since feedback use group using only the tablet did not perform exercises with the feedback system.

<sup>c</sup>UPS: unipedal stance.

<sup>d</sup>CRT: chair rise test.

<sup>e</sup>YBT: Y-balance test.

<sup>f</sup>HS: hurdle step.

## Effect of Exercise Modules on Balance and Leg Strength

### Overview

The statistical results are presented in [Table 3](#) without considering use adherence and in [Table 4](#) considering use adherence.

**Table 3.** Statistical test results comparing the change of functional assessment results between baseline and feedback use groups without considering use adherence (N=83).

Feedback use group	Baseline		Using only the tablet		Using the tablet and the feedback system		Using only the feedback system		P value	Effect size
	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)		
ΔUnipedal stance maximum	82 (99)	2.3 (8.5) <sup>a</sup>	27 (33)	4.4 (22.6)	31 (37)	1.9 (11.6)	4 (5)	8.8 (14.4)	.62	<0.06
ΔUnipedal stance left leg maximum	82 (99)	4.2 (13.4) <sup>a</sup>	27 (33)	1.9 (27.4)	31 (37)	5.6 (19.4)	4 (5)	6.3 (39.3)	.94	<0.06
ΔUnipedal stance right leg maximum	82 (99)	3.2 (10.4) <sup>a</sup>	27 (33)	7.2 (28.5)	31 (37)	2.0 (11.7)	4 (5)	11.3 (13.1)	.23	<0.06
Δ30-second chair rise test	80 (96)	0.4 (2.7) <sup>a</sup>	26 (31)	5.8 (6.3) <sup>a,b</sup>	31 (37)	2.3 (5.9) <sup>a</sup>	5 (6)	6.3 (3.6) <sup>b</sup>	<.001 <sup>c</sup>	0.18 <sup>d</sup>
ΔY-balance test left leg	71 (86)	0.7 (9.6) <sup>a</sup>	24 (29)	7.1 (13.7) <sup>a</sup>	31 (37)	-1.4 (14.1) <sup>e</sup>	4 (5)	15.2 (12.6) <sup>b,f</sup>	.01 <sup>c</sup>	0.07 <sup>g</sup>
ΔY-balance test right leg	71 (86)	1.81 (10.1) <sup>a</sup>	24 (29)	6.2 (12.2) <sup>a</sup>	31 (37)	-2.0 (10.5)	4 (5)	10.2 (12.6)	.04 <sup>c</sup>	<0.06
ΔY-balance test symmetry	71 (86)	-0.7 (7.1)	24 (29)	-1.2 (3.3)	31 (37)	-0.3 (6.0)	4 (5)	-0.4 (2.8)	.42	<0.06
ΔHurdle step left leg	79 (95)	0.0 (0.7)	25 (30)	-0.2 (0.9)	31 (37)	-0.3 (0.8) <sup>a</sup>	4 (5)	0.0 (1.4)	.29	<0.06
ΔHurdle step right leg	79 (95)	0.1 (0.7)	25 (30)	-0.1 (0.8)	31 (37)	-0.3 (0.8) <sup>a</sup>	4 (5)	0.0 (1.4)	.11	<0.06

<sup>a</sup>Significance within groups.

<sup>b</sup>Significant in comparison with the baseline group.

<sup>c</sup>Significance level  $\alpha < .05$ .

<sup>d</sup>Large effect  $\geq 0.14$ .

<sup>e</sup>Significant in comparison with the feedback use group using only the tablet.

<sup>f</sup>Significant in comparison with the feedback use group using the tablet and the feedback system.

<sup>g</sup>Moderate effect between 0.06 and 0.14.

**Table 4.** Statistical test results comparing the change of functional assessment results between baseline and feedback use groups considering use adherence (N=83).

Feedback use group	Baseline		Using only the tablet		Using the tablet and the feedback system		Using only the feedback system		P value	Effect size
	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)	Values, n (%)	Values, mean (SD)		
ΔUnipedal stance maximum	82 (99)	3.0 (8.5) <sup>a</sup>	9 (11)	-10.9 (20.7)	22 (27)	0.7 (8.5)	1 (1)	0.0	.17	<0.06
ΔUnipedal stance left leg maximum	82 (99)	4.8 (13.5) <sup>a</sup>	9 (11)	-15.9 (22.8)	22 (27)	3.3 (18.2)	1 (1)	-47.0	.02 <sup>b</sup>	0.07 <sup>c</sup>
ΔUnipedal stance right leg maximum	82 (99)	3.9 (10.4) <sup>a</sup>	9 (11)	-4.1 (35.0)	22 (27)	0.8 (8.6)	1 (1)	0.0	.90	<0.06
Δ30-second chair rise test	80 (96)	0.4 (2.7)	9 (11)	3.8 (4.5) <sup>a</sup>	22 (27)	2.8 (6.2) <sup>a</sup>	1 (1)	6.5	.01 <sup>b</sup>	0.07 <sup>c</sup>
ΔY-balance test left leg	71 (86)	0.7 (9.6) <sup>a</sup>	7 (8)	2.6 (11.1)	22 (27)	-1.2 (15.8)	1 (1)	2.7	.33	<0.06
ΔY-balance test right leg	71 (86)	1.8 (10.1)	7 (8)	2.9 (10.2)	22 (27)	-2.7 (12.0)	1 (1)	-5.0	.20	<0.06
ΔY-balance test symmetry	71 (86)	-0.7 (7.1)	7 (8)	-0.2 (3.7)	22 (27)	-1.2 (6.0)	1 (1)	-1.9	.91	<0.06
ΔHurdle step left leg	79 (95)	0.0 (0.7)	8 (10)	-0.1 (0.8)	22 (27)	-0.4 (0.8)	1 (1)	0.0	.30	<0.06
ΔHurdle step right leg	79 (95)	0.1 (0.7)	8 (10)	-0.3 (0.7)	22 (27)	-0.3 (0.7)	1 (1)	0.0	.17	<0.06

<sup>a</sup>Significance within groups.

<sup>b</sup>Significance level  $\alpha < .05$ .

<sup>c</sup>Moderate effect between 0.06 and 0.14.

### Test Participation

The participants were able to withdraw themselves from performing 1 or all of the tests at any time. Several participants did not want to conduct single tests on site, although they would have been able to do so. The data availability of the functional performance assessments for each participant (ie, their test participation) varied depending on time point and type of test. The number of functional performance assessments conducted increased from basis to preintervention measurements: the number of participants who conducted the 30-second CRT increased from 96% (80/83) to 99% (82/83), YBT participation increased from 86% (71/83) to 96% (80/83), and HS test participation increased from 95% (79/83) to 98% (81/83). UPS test participation remained almost steady at approximately 99% (82/83) between baseline and preintervention data availability. In contrast, more preintervention than postintervention data on functional performance were available: the UPS test and CRT participation dropped from 99% (82/83) to 75% (62/83), the YBT participation dropped from 96% (80/83) to 71% (59/83), and the HS test data availability decreased from 98% (81/83) to 72% (60/83).

The reasons for the differences in participation were mainly scheduling problems because of the tight test schedule. If a person was late and, therefore, the test was only possible for a shorter time, there was no possibility to reschedule. Furthermore, the participants had the option of refusing to take a test. At the end of the intervention, not all study participants were convinced that they had benefited from the tests as the system simply had to be returned. Another less common reason was the weather. Temperatures were not always optimal for testing; thus, the participants were concerned about circulatory problems. Again,

because of the tight schedule, no alternative date could be offered by the test team.

### Not Considering Use Adherence

Significant improvements and interaction effects were mainly found for leg strength and dynamic balance when comparing the feedback use groups and the baseline difference.

The average static balance changes (UPS) were not different between the groups. Within the baseline group, the static balance (UPS) improved between basis and preintervention ( $P < .001$ ).

The leg strength improvement of the 3 feedback use groups was significant in comparison with the baseline difference ( $P < .001$ ), with a large effect size of 0.179. Pairwise comparisons determined that participants using either of the 2 exercise modules improved in leg strength compared with the baseline difference ( $P < .001$  for feedback use group using only the tablet, and  $P = .01$  for the feedback use group using only the feedback system). Furthermore, within the baseline difference ( $P = .03$ ), the feedback use group using only the tablet ( $P < .001$ ), and the feedback use group using the tablet and the feedback system ( $P = .02$ ), the average leg strength comparing pre- and postintervention values was significant.

The dynamic balance improvement of the left leg (YBT left leg composite score) for the 2 feedback use groups (1) using only the tablet and (2) using only the feedback system was significant compared with the baseline difference, whereas it slightly decreased for feedback use group using the tablet and the feedback system in comparison with the baseline ( $P = .006$ ; moderate effect size of 0.074). The pairwise comparison showed an improvement of dynamic balance change for the feedback use group using only the feedback system related to the baseline difference ( $P = .03$ ) and to the feedback use group using the tablet

and the feedback system ( $P=.03$ ). The slight dynamic balance decrease in the left leg of the feedback use group using the tablet and the feedback system was significant compared with that of the feedback use group using only the tablet ( $P=.03$ ).

Although the dynamic balance differences of the right leg were significant between the groups ( $P=.04$ ; small effect size of 0.043), the pairwise comparisons were not significant. Dynamic balance comparing pre- and postintervention measures improved within the feedback use group using only the tablet (YBT of the right and the left legs,  $P=.02$ ). Within the baseline difference, the dynamic balance performance improved comparing baseline and preintervention values (YBT of left and right leg,  $P=.01$ ).

Comparing the pre- and postintervention values within the feedback use group using the tablet and the feedback system, the movement quality slightly decreased on average (HS left leg  $P=.04$  and HS right leg  $P=.03$ ). No further significance was observed.

### Considering Use Adherence

When including only users who used the exercise modules at the WHO-recommended training frequency (Table 4), the baseline group showed significant improvement in static balance (UPS,  $P<.001$ ) and dynamic balance of the left leg (YBT left leg,  $P=.005$ ). Another within-group improvement in leg strength was identified in the feedback use group using only the tablet ( $P=.02$ ) and in the feedback use group using the tablet and the feedback system ( $P=.04$ ). The feedback use groups (using only the tablet, using the tablet and the feedback system, using only the feedback system) achieved, on average, more chair rises in 30 seconds than the baseline group ( $P=.01$ ; moderate effect size of 0.073). In addition, the static balance of the left leg (UPS left leg maximum) of the feedback use group using the tablet and the feedback system and the baseline group improved in comparison with the feedback use groups using only the tablet and only the feedback system ( $P=.02$ ; moderate effect size of 0.065). Pairwise comparisons did not reveal any further significance.

## Discussion

### Principal Findings

The purpose of this study was to assess the influence of digital exercise modules of a multimodal solution on balance and leg strength. As 2 digital exercise modules were administered within the study, use adherence to these 2 modules was considered to identify the users who met the recommended exercise frequency according to the WHO. To the best of our knowledge, this is the first study to evaluate the influence of single modules of a multimodal AAL solution on balance and leg strength considering use adherence. When considering the use adherence, a tendency toward a positive influence on leg strength was found for participants using the tablet-based or additionally using the feedback system-based exercise module but not between the feedback use groups. Without considering the use adherence to the digital home training, more positive effects were indicated; however, these are more likely to be induced by the other modules of the multimodal AAL solution or external influences than the digital exercise modules.

Therefore, use adherence should be considered in the future for functional performance assessments of multimodal physical activity-promoting applications.

In this study, use adherence indicated that 52% (43/83) of the digital home training users were able to meet the WHO-recommended training frequency. This is almost twice the number of Austrian men and women aged 45 to 64 years who reported reaching the WHO-recommended frequency for muscle strengthening in 2019 (27.1% and 26.7%, respectively) [32].

Nevertheless, 1 cluster of user types was unable to comply with the training frequency recommended by the WHO. Previous research investigated the adherence to the modules of a multidomain lifestyle training, including cognitive training, nutrition, and exercise, provided on a tablet [33]. Although their participants mainly used the cognitive training module, they reported that the exercise module lacked diversity, challenges, and progression. The reasons in the fit4AAL study require further investigation; for example, if the study participants who did not meet the WHO-recommended training frequency used other modules more than the exercise modules and why.

However, the additional use of the feedback system tended to support the users to maintain the WHO-recommended training frequency and showed higher exercise doses. This is in alignment with the findings of the review by Brickwood et al [34], who showed that empowering technologies such as the commonly available activity trackers positively influenced physical activity participation.

The administered training program has already shown a positive influence on functional performance in older women in the first trial phase [19]. Nevertheless, the extent to which the choice of technology influences these effects has not yet been investigated. Exercise programs at home have already been proven to improve balance and reduce fall reduction rates in older adults [35-37].

Although the results of the analysis without considering use adherence confirmed the improvement in balance and leg strength, a different picture emerges when considering the use adherence: the improvement of leg strength remained within the feedback use groups using only the tablet and using the tablet and the feedback system. Improvements for the static balance only remained within the baseline group. Although the exercise modules improved the leg strength, a mean decrease in balance performance was observed. The decrease in static balance in the feedback use group using only the tablet was remarkable, whereas, in the feedback use group using the tablet and the feedback system, the static balance of the left leg remained or even improved.

This could be explained by the fact that the balance of the left leg could be increased by the intervention compared with the more likely dominant right leg. Moreover, digital exercise modules relying on tablet-based solutions without any training feedback might, on average, negatively influence the static balance of the participants. This strengthens the findings on slight reductions of static balance capabilities of regular app-using test groups in the course of multimonth studies [38]. Comparing the average pre- and postintervention assessment

differences, all feedback use groups almost maintained their functional performance levels. Pairwise comparisons did not show any significance when considering the use adherence.

A possible explanation for the missing pairwise effects between the feedback use groups and the baseline could be that the functional performance test results of the baseline and preintervention assessments of the investigated participants were already within the age-appropriate norm values. Our sample of older adults was anything but fragile: the users were within and even beyond the norm values of their age group. For example, the minimum mean values of 15.3 (SD 3.5) chair rises (Table 1) already exceeded the norm values of 13.5 (SD 3.5) chair rises for the 30-second CRT for women aged 65 to 69 years [39]. In addition, the minimum range of 49.7 (SD 17.7) seconds of the UPS (Table 1) exceeded the age-appropriate norm value of 32.1 (SD 16.2) seconds [24]. Hence, the study participants were exceptionally fit. A focus on maintenance or even improvement of the functional performance values would require updated norm values of the age group and the region they are coming from. In Austria, the difference in complying with the WHO recommendations on physical activity varies between provinces [32]. Future research should investigate sample groups with different fitness levels in different age groups.

Moreover, a study design with groups using both exercise modules versus using only one of them with sufficient group size would be beneficial. Although a positive influence within the feedback use groups in leg strength was identified, this should be verified with a larger sample size separated into exercise module group, nondigital intervention group, and control group for more sufficient effect sizes [37,40].

In summary, the additional use of the feedback system-based exercise module might not yet verify the improvement of the functional performance in leg strength and balance of older adults who were trained at least twice a week. Nevertheless, the digital exercise modules of a multimodal AAL solution showed a tendency to positively maintain and influence the already exceptionally high functional performance in older adults. Moreover, the use of additional use of empowering technologies could support users to achieve their training goals, keeping the known challenges such as usability and technology acceptance in mind.

As 2 digital exercise modules within a multimodal AAL solution were used within an 18-week intervention, this study is the first attempt to investigate functional performance outcomes considering the use adherence of single modules, except the entire AAL solution, in older adults. The main recommendations for future field trials, in particular, AAL field studies, is to focus on modules and consider or rather nudge the voluntary use. Our intervention group was able to use all modules instead of only particular modules. This intended and resulted in maintaining high adherence and engagement with

the system. As practical implications for further field studies aiming to investigate the influence of particular modules on selected markers of functional performance of the participants, it would be beneficial to use already evaluated modules in the AAL study or to test newly developed exercise modules in a prefield trial with friendly users independent from other modules. Hence, two main challenges for the analysis procedure were identified: (1) the multimodularity of solutions, which requires user type clustering to identify whether and how often the particular modules were used, and (2) the selection of specific intervention measures describing the influence of particular modules.

### Limitations

A limitation of this study is that no paper-based interventions were evaluated. Owing to the separation into feedback use groups, the sample size for the statistical analysis decreased; thus, the statistical power and generalizability were reduced. In future studies, more users should be included to reduce the risk of small sample sizes when clustering the users. Furthermore, additional motivation strategies to keep users engaged with the system, as well as the functional performance tests, have to be considered. Although the recruitment targeted a balanced study sample, more women than men applied for the study. For generalizability, not only a larger sample size but also a sex-balanced sample size would be beneficial for analysis by gender with sufficient statistical power. Whether the system or study attracted fewer men than women was not investigated.

### Conclusions

Although there are variations in the use of the digital exercise modules, the additional use of a feedback system in the multimodal AAL solution of the fit4AAL project positively influences use adherence and improvements in leg strength within feedback use groups. Thus, involving various empowering technologies to keep people engaged in their unsupervised digital home training can be recommended for active aging. The feedback system in this study integrates a split-screen view and posture tracking for repetition counting and exercise monitoring for selected exercises. In the future, more movement quality-monitoring functionalities, integrating further trackers or even on-body sensor networks of, for example, smart textiles and smartwatches, could be considered. To further justify the positive influence of the additional feedback system-based exercise module on functional performance, a different study design must be considered with, for example, different age groups, because of different technology savvy and functional performance levels. The participants of this study were exceptionally fit compared with the norm values of their age groups. Whether the digital exercise modules in multimodal AAL solutions can generate more benefit in establishing a home training routine in older adults has to be clarified in future longitudinal studies.

## Acknowledgments

This research was funded by the Austrian Federal Ministry for Transport, Innovation, and Technology under the research program benefit of the cooperative project fit4AAL (“Fit in einen neuen Lebensabschnitt mit neuen Technologien—AAL-Testregion Salzburg/Wien”; grant number 862035 by the Austrian Research Promotion Agency [Österreichische Forschungsförderungsgesellschaft]). For the development of the exercise modules and the conduction of the field trial, Salzburg Research, MyBodyCoach, and the Department of Sport and Exercise Science of the Paris Lodron Universität Salzburg want to thank their additional consortium partners: bit media e-solutions GmbH, SMART ASSETS Development GmbH, 50plus GmbH, Salzburg AG, Care Consulting (Sonja Schiff), and the Wirtschaftsuniversität Wien (Forschungsinstitut für Altersökonomie). The authors would also like to thank all the study participants in Salzburg and Vienna for their participation.

## Authors' Contributions

All authors are qualified for authorship of the article following the criteria of the International Committee of Medical Journal Editors guidelines. VV drafted the manuscript with the help of CK, SJ, SRD, HS, and TS. All authors contributed to the critical revision of the draft and approved the final version of the manuscript.

## Conflicts of Interest

None declared.

## References

1. Healthy ageing and functional ability. World Health Organization. 2020 Oct 26. URL: <https://www.who.int/philippines/news/q-a-detail/healthy-ageing-and-functional-ability> [accessed 2021-12-12]
2. Ohrnberger J, Fichera E, Sutton M. The relationship between physical and mental health: a mediation analysis. *Soc Sci Med* 2017 Dec;195:42-49 [FREE Full text] [doi: [10.1016/j.socscimed.2017.11.008](https://doi.org/10.1016/j.socscimed.2017.11.008)] [Medline: [29132081](https://pubmed.ncbi.nlm.nih.gov/29132081/)]
3. World Health Organization. Promoting physical activity in the health sector: current status and success stories from the European Union Member States of the WHO European Region. World Health Organization Regional Office for Europe. 2018. URL: <https://apps.who.int/iris/handle/10665/345137> [accessed 2021-12-12]
4. Vichitvanichphong S, Talaei-Khoei A, Kerr D. Elderly's perception about the value of assistive technologies for their daily living: impacting factors and theoretical support. In: Proceedings of the 50th Hawaii International Conference on System Sciences. 2017 Presented at: 50th Hawaii International Conference on System Sciences; Jan 4-7, 2017; Village, Hawaii, USA. [doi: [10.24251/hicss.2017.445](https://doi.org/10.24251/hicss.2017.445)]
5. Room J, Hannink E, Dawes H, Barker K. What interventions are used to improve exercise adherence in older people and what behavioural techniques are they based on? A systematic review. *BMJ Open* 2017 Dec 14;7(12):e019221 [FREE Full text] [doi: [10.1136/bmjopen-2017-019221](https://doi.org/10.1136/bmjopen-2017-019221)] [Medline: [29247111](https://pubmed.ncbi.nlm.nih.gov/29247111/)]
6. Gschwind YJ, Eichberg S, Ejupi A, de Rosario H, Kroll M, Marston HR, et al. ICT-based system to predict and prevent falls (iStoppFalls): results from an international multicenter randomized controlled trial. *Eur Rev Aging Phys Act* 2015;12:10 [FREE Full text] [doi: [10.1186/s11556-015-0155-6](https://doi.org/10.1186/s11556-015-0155-6)] [Medline: [26865874](https://pubmed.ncbi.nlm.nih.gov/26865874/)]
7. Van Grootven B, van Achterberg T. The European Union's ambient and assisted living joint programme: an evaluation of its impact on population health and well-being. *Health Informatics J* 2019 Mar;25(1):27-40 [FREE Full text] [doi: [10.1177/1460458216683535](https://doi.org/10.1177/1460458216683535)] [Medline: [28006968](https://pubmed.ncbi.nlm.nih.gov/28006968/)]
8. Wanner M, Martin-Diener E, Braun-Fahrländer C, Bauer G, Martin BW. Effectiveness of active-online, an individually tailored physical activity intervention, in a real-life setting: randomized controlled trial. *J Med Internet Res* 2009 Jul 28;11(3):e23 [FREE Full text] [doi: [10.2196/jmir.1179](https://doi.org/10.2196/jmir.1179)] [Medline: [19666456](https://pubmed.ncbi.nlm.nih.gov/19666456/)]
9. Vandelanotte C, Duncan MJ, Kolt GS, Caperchione CM, Savage TN, Van Itallie A, et al. More real-world trials are needed to establish if web-based physical activity interventions are effective. *Br J Sports Med* 2019 Dec 03;53(24):1553-1554. [doi: [10.1136/bjsports-2018-099437](https://doi.org/10.1136/bjsports-2018-099437)] [Medline: [29970409](https://pubmed.ncbi.nlm.nih.gov/29970409/)]
10. Dasgupta D, Chaudhry B, Koh E, Chawla NV. A survey of tablet applications for promoting successful aging in older adults. *IEEE Access* 2016;4:9005-9017. [doi: [10.1109/ACCESS.2016.2632818](https://doi.org/10.1109/ACCESS.2016.2632818)]
11. Global Recommendations on Physical Activity for Health. Geneva: World Health Organization; Jan 1, 2010.
12. Withall J, Jago R, Fox KR. Why some do but most don't. Barriers and enablers to engaging low-income groups in physical activity programmes: a mixed methods study. *BMC Public Health* 2011 Jun 28;11(1):507 [FREE Full text] [doi: [10.1186/1471-2458-11-507](https://doi.org/10.1186/1471-2458-11-507)] [Medline: [21711514](https://pubmed.ncbi.nlm.nih.gov/21711514/)]
13. Mendoza-Vasconez AS, Linke S, Muñoz M, Pekmezi D, Ainsworth C, Cano M, et al. Promoting physical activity among underserved populations. *Current Sports Med Report* 2016;15(4):290-297. [doi: [10.1249/jsr.0000000000000276](https://doi.org/10.1249/jsr.0000000000000276)]
14. Vseteckova J, Deepak-Gopinath M, Borgstrom E, Holland C, Draper J, Pappas Y, et al. Barriers and facilitators to adherence to group exercise in institutionalized older people living with dementia: a systematic review. *Eur Rev Aging Phys Act* 2018 Nov 9;15(1):11 [FREE Full text] [doi: [10.1186/s11556-018-0200-3](https://doi.org/10.1186/s11556-018-0200-3)] [Medline: [30455778](https://pubmed.ncbi.nlm.nih.gov/30455778/)]
15. Trukeschitz B, Schneider C, Ring-Dimitriou S. Smartes Betreutes Wohnen Nutzung, Systemakzeptanz und Wirkungen von „meinZentrAAL“. Norderstedt, Deutschland: Book on Demand; 2017.

16. Schneider C, Trukeschitz B, Rieser H. Measuring the use of the active and assisted living prototype CARIMO for home care service users: evaluation framework and results. *Applied Sci* 2019 Dec 19;10(1):38. [doi: [10.3390/app10010038](https://doi.org/10.3390/app10010038)]
17. Belleville S, Cuesta M, Bieler-Aeschlimann M, Giacomino K, Widmer A, Mittaz Hager AG, et al. Rationale and protocol of the StayFitLonger study: a multicentre trial to measure efficacy and adherence of a home-based computerised multidomain intervention in healthy older adults. *BMC Geriatr* 2020 Aug 28;20(1):315 [FREE Full text] [doi: [10.1186/s12877-020-01709-2](https://doi.org/10.1186/s12877-020-01709-2)] [Medline: [32859156](https://pubmed.ncbi.nlm.nih.gov/32859156/)]
18. Trukeschitz B, Blüher M, Schneider C, Jungreitmayr S, Eisenberg S. „Fit-mit-ILSE“ Feldtest: Design, Rekrutierung und Übersicht über die TeilnehmerInnen zu Beginn des Feldtests. *fit4AAL*. 2019. URL: [https://www.salzburgresearch.at/wp-content/uploads/2021/01/fit4AAL-D14\\_Feldtestdesign.pdf](https://www.salzburgresearch.at/wp-content/uploads/2021/01/fit4AAL-D14_Feldtestdesign.pdf) [accessed 2021-11-07]
19. Jungreitmayr S, Kranzinger C, Venek V, Ring-Dimitriou S. Effects of an app-based physical exercise program on selected parameters of physical fitness of females in retirement: a randomized controlled trial. *Front Physiol* 2022 Mar 4;13:821773 [FREE Full text] [doi: [10.3389/fphys.2022.821773](https://doi.org/10.3389/fphys.2022.821773)] [Medline: [35317213](https://pubmed.ncbi.nlm.nih.gov/35317213/)]
20. Jungreitmayr S. Auswirkungen von ILSE auf die funktionale Fitness. *fit4AAL*. 2021. URL: [https://www.salzburgresearch.at/wp-content/uploads/2021/03/fit4AAL-Evalu-Report-SJ\\_update-feb2021.pdf](https://www.salzburgresearch.at/wp-content/uploads/2021/03/fit4AAL-Evalu-Report-SJ_update-feb2021.pdf) [accessed 2022-01-08]
21. Jungreitmayr S. Krafttraining im Alter. *Sportphysio* 2018 Sep 07;06(04):161-170. [doi: [10.1055/a-0642-8120](https://doi.org/10.1055/a-0642-8120)]
22. Venek V, Kremser W, Stöggel T. Towards a live feedback training system: interchangeability of Orbbeo Persee and Microsoft Kinect for exercise monitoring. *Designs* 2021 Apr 15;5(2):30. [doi: [10.3390/designs5020030](https://doi.org/10.3390/designs5020030)]
23. Neuwirth C, Venek V, Rieser H, Maringer V. Vergleich der ILSE-App Nutzung in den beiden Feldtestphasen. *fit4AAL*. 2020. URL: [https://www.salzburgresearch.at/wp-content/uploads/2021/01/fit4AAL-Nutzungsanalyse-ILSE-FT2\\_final\\_Websiteversion.pdf](https://www.salzburgresearch.at/wp-content/uploads/2021/01/fit4AAL-Nutzungsanalyse-ILSE-FT2_final_Websiteversion.pdf) [accessed 2021-11-18]
24. 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](https://doi.org/10.1519/00139143-200704000-00003)] [Medline: [19839175](https://pubmed.ncbi.nlm.nih.gov/19839175/)]
25. Wilson JJ, McMullan I, Blackburn NE, Skjødt M, Caserotti P, Giné-Garriga M, et al. Associations of sedentary behavior bouts with community-dwelling older adults' physical function. *Scand J Med Sci Sports* 2021 Jan;31(1):153-162. [doi: [10.1111/sms.13827](https://doi.org/10.1111/sms.13827)] [Medline: [32945566](https://pubmed.ncbi.nlm.nih.gov/32945566/)]
26. Bohannon R. Measurement of sit-to-stand among older adults. *Top Geriatr Rehabil* 2012;28(1):11-16. [doi: [10.1097/TGR.0b013e31823415fa](https://doi.org/10.1097/TGR.0b013e31823415fa)]
27. Sipe C, Ramey K, Plisky P, Taylor J. Y-balance test: a valid and reliable assessment in older adults. *J Aging Phys Act* 2019 Sep 01;27(5):663-669. [doi: [10.1123/japa.2018-0330](https://doi.org/10.1123/japa.2018-0330)] [Medline: [30676192](https://pubmed.ncbi.nlm.nih.gov/30676192/)]
28. Cook G, Burton L, Hoogenboom BJ, Voight M. Functional movement screening: the use of fundamental movements as an assessment of function - part 1. *Int J Sports Phys Ther* 2014 May;9(3):396-409 [FREE Full text] [Medline: [24944860](https://pubmed.ncbi.nlm.nih.gov/24944860/)]
29. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
30. Kassambara A. rstatix: pipe-friendly framework for basic statistical tests. CRAN. 2020. URL: <https://cran.r-project.org/package=rstatix> [accessed 2022-01-18]
31. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Series B (Methodological)* 2018 Dec 05;57(1):289-300. [doi: [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)]
32. Österreichische Gesundheitsbefragung 2019: Hauptergebnisse des Austrian Health Interview Survey (ATHIS) und methodische Dokumentation. *Statistik Austria*. 2019. URL: [http://www.statistik.at/web\\_de/services/publikationen/4/index.html?includePage=detailedView&sectionName=Gesundheit&pubId=794](http://www.statistik.at/web_de/services/publikationen/4/index.html?includePage=detailedView&sectionName=Gesundheit&pubId=794) [accessed 2022-01-18]
33. de Souto Barreto P, Pothier K, Soriano G, Lussier M, Bherer L, Guyonnet S, et al. A web-based multidomain lifestyle intervention for older adults: the eMIND randomized controlled trial. *J Prev Alzheimers Dis* 2021;8(2):142-150 [FREE Full text] [doi: [10.14283/jpad.2020.70](https://doi.org/10.14283/jpad.2020.70)] [Medline: [33569560](https://pubmed.ncbi.nlm.nih.gov/33569560/)]
34. 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](https://doi.org/10.2196/11819)] [Medline: [30977740](https://pubmed.ncbi.nlm.nih.gov/30977740/)]
35. Papalia GF, Papalia R, Diaz Balzani LA, Torre G, Zampogna B, Vasta S, et al. The effects of physical exercise on balance and prevention of falls in older people: a systematic review and meta-analysis. *J Clin Med* 2020 Aug 11;9(8):2595 [FREE Full text] [doi: [10.3390/jcm9082595](https://doi.org/10.3390/jcm9082595)] [Medline: [32796528](https://pubmed.ncbi.nlm.nih.gov/32796528/)]
36. Gschwind YJ, Schoene D, Lord SR, Ejupi A, Valenzuela T, Aal K, et al. The effect of sensor-based exercise at home on functional performance associated with fall risk in older people - a comparison of two exergame interventions. *Eur Rev Aging Phys Act* 2015 Nov 30;12(1):11 [FREE Full text] [doi: [10.1186/s11556-015-0156-5](https://doi.org/10.1186/s11556-015-0156-5)] [Medline: [26865875](https://pubmed.ncbi.nlm.nih.gov/26865875/)]
37. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. *Prev Med* 2018 Mar;108:93-110. [doi: [10.1016/j.ypmed.2017.12.026](https://doi.org/10.1016/j.ypmed.2017.12.026)] [Medline: [29289643](https://pubmed.ncbi.nlm.nih.gov/29289643/)]
38. Jungreitmayr S, Ring-Dimitriou S, Trukeschitz B, Eisenberg S, Schneider C. Effects of an information and communication technology-based fitness program on strength and balance in female home care service users. *Int J Environ Res Public Health* 2021 Jul 28;18(15):7955 [FREE Full text] [doi: [10.3390/ijerph18157955](https://doi.org/10.3390/ijerph18157955)] [Medline: [34360249](https://pubmed.ncbi.nlm.nih.gov/34360249/)]
39. Rikli R, Jones C. Functional fitness normative scores for community-residing older adults, ages 60-94. *J Aging Phys Act* 1999;7(2):162-182. [doi: [10.1123/japa.7.2.162](https://doi.org/10.1123/japa.7.2.162)]



40. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: a systematic review and meta-analysis. *Exp Gerontol* 2019 Jun;120:68-87. [doi: [10.1016/j.exger.2019.02.020](https://doi.org/10.1016/j.exger.2019.02.020)] [Medline: [30836130](https://pubmed.ncbi.nlm.nih.gov/30836130/)]

## Abbreviations

**AAL:** Active and Assisted Living

**CRT:** chair rise test

**HS:** hurdle step

**ICT:** Information and Communication Technology

**UPS:** unipedal stance

**WHO:** World Health Organization

**YBT:** Y-balance test

*Edited by A Mavragani; submitted 26.01.22; peer-reviewed by A Garcia, S Bonn; comments to author 29.06.22; revised version received 19.07.22; accepted 09.08.22; published 19.09.22.*

*Please cite as:*

*Venek V, Kranzinger C, Jungreitmayr S, Ring-Dimitriou S, Schwameder H, Stöggel T*

*Influence of 2 Digital Exercise Modules of a Multimodular System on Balance and Leg Strength Under Consideration of Use Adherence: Prospective Cohort Study*

*JMIR Form Res* 2022;6(9):e36805

URL: <https://formative.jmir.org/2022/9/e36805>

doi: [10.2196/36805](https://doi.org/10.2196/36805)

PMID: [36121691](https://pubmed.ncbi.nlm.nih.gov/36121691/)

©Verena Venek, Christina Kranzinger, Sonja Jungreitmayr, Susanne Ring-Dimitriou, Hermann Schwameder, Thomas Stöggel. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 19.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# A Web-Based Positive Psychology App for Patients With Bipolar Disorder: Development Study

Bart Geerling<sup>1,2</sup>, MANP; Saskia M Kelders<sup>1,3</sup>, PhD; Anja W M M Stevens<sup>2</sup>, MD; Ralph W Kupka<sup>4</sup>, Prof Dr; Ernst T Bohlmeijer<sup>1</sup>, Prof Dr

<sup>1</sup>Department of Psychology, Health and Technology, Centre for eHealth and Wellbeing Research, University of Twente, Enschede, Netherlands

<sup>2</sup>Centre for Bipolar Disorder, Dimence Mental Health Institute, Deventer, Netherlands

<sup>3</sup>Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa

<sup>4</sup>Department of Psychiatry, Mental Health Program, Amsterdam Public Health, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, Netherlands

**Corresponding Author:**

Bart Geerling, MANP

Department of Psychology, Health and Technology

Centre for eHealth and Wellbeing Research

University of Twente

Drienerlolaan 5

Enschede, 7522 NB

Netherlands

Phone: 31 0622274351

Email: [b.geerling@dimence.nl](mailto:b.geerling@dimence.nl)

## Abstract

**Background:** Patients with bipolar disorder (BD) report lower quality of life and lower levels of well-being than the general population. Despite the growing availability of psychotherapeutic and self-management interventions, important unmet needs remain. These unmet needs are closely linked to positive psychology domains. Although a growing number of studies have evaluated the impact of positive psychology interventions (PPIs) on patients with severe mental illness in general, only few have addressed the application of positive psychology for BD.

**Objective:** This study aimed to gain insight into the opinions of patients with BD and health care professionals about (web-based) PPIs for BD and to develop and pilot-test an app containing PPIs specifically designed for patients with BD.

**Methods:** The study was conducted in accordance with the Center for eHealth and Disease Management road map principles and incorporated cocreation and designing for implementation. Data were collected using focus group discussions, questionnaires, rapid prototyping, and web-based feedback on a prototype from the participants. In total, 3 focus groups were conducted with 62% (8/13) of patients with BD and 38% (5/13) of professionals. The collected data were used to develop a smartphone app containing short PPIs. The content was based on PPIs for which a solid base of evidence is available. Finally, a pilot test was conducted to test the app.

**Results:** Focus groups revealed that PPIs as part of the current BD treatment can potentially meet the following needs: offering hope, increasing self-esteem, expressing feelings, acceptance, and preventing social isolation. Some patients expressed concern that PPIs may provoke a manic or hypomanic episode by increasing positive affect. The pilot of the app showed that the PPIs are moderately to highly valued by the participants. There were no adverse effects such as increase in manic or hypomanic symptoms.

**Conclusions:** With the systematic use of user involvement (patients and professionals) in all steps of the development process, we were able to create an app that can potentially fulfill some of the current unmet needs in the treatment of BD. We reached consensus among consumers and professionals about the potential benefits of PPIs to address the unmet needs of patients with BD. The use of PPI for BD is intriguing and can be usefully explored in further studies. We emphasize that more evaluation studies (quantitative and qualitative) that are focused on the effect of PPIs in the treatment of BD should be conducted. In addition, to establish the working mechanisms in BD, explorative, qualitative, designed studies are required to reveal whether PPIs can address unmet needs in BD.

(*JMIR Form Res* 2022;6(9):e39476) doi:[10.2196/39476](https://doi.org/10.2196/39476)

**KEYWORDS**

bipolar disorder; positive psychology; cocreation; mobile health; mHealth; web-based; psychology; bipolar; intervention; quality of life; mental illness; pilot; self-esteem; acceptance; social isolation; manic episode; manic; self-help; positive; mobile phone

## Introduction

Bipolar disorder (BD) is defined as a chronic mental illness with recurrent mood episodes, with manic, hypomanic, and depressive episodes alternating with euthymic periods. The illness mostly begins in adolescence and young adulthood [1]. It is estimated that BD I and II disorders occur in 2% of the world's population, and another estimated 2% has a subthreshold BD [2]. In the Netherlands, the prevalence of BD in the adult population is 1.3% [3].

Patients and their significant others face significant burden when confronted with BD. Owing to the early onset, severity, and chronicity, BD is a potentially disabling illness [4-6]. Even during euthymic periods between episodes, the illness may lead to impairment and significant burden [7]. Factors contributing to the burden are persistent subsyndromal mood symptoms, stigmatization, cognitive impairment, comorbid conditions, and side effects of pharmacotherapy [7]. The burden of BD may increase over the longitudinal course of the disease. Illness progression has been described using various staging models for BD [8], and it is expected that the burden of illness is more prominent in the later stages with multiple recurrences or persistent unremitting illness [9]. Persistence of mood symptoms between episodes is a significant predictor of depression and functional impairment [10]. The level of (subsyndromal) depressive symptoms correlates positively with the degree of functional impairment [11] and, therefore, with a high burden and low quality of life. Despite the growing availability of psychotherapeutic and self-management interventions, important unmet needs remain, including those that are not directly related to mood symptoms in patients with BD.

Previous studies have established that there are major unmet needs in the management of BD [12-17]. Needs are defined as what people "desire to receive from healthcare services to improve overall health" [18]. The most common needs during depression and mania or hypomania can be summarized as encouragement to seek effective (pharmacological) treatment to reduce symptoms. During remission and subsyndromal episodes, there is a need for treatment that prevents future episodes and a need for easily available psychosocial interventions [13,16].

Most common needs are satisfied to some degree during regular treatment. Some studies have categorized additional unmet needs addressing social and psychological functioning using questionnaires (Need for Care Questionnaire or self-developed questionnaires) [14,15]. In social functioning, support with loneliness, grief counseling, acceptance, social isolation, and coping with others are frequently mentioned [14,15,17]. Hope, expressing feelings, and increasing self-confidence are mentioned as unmet needs in psychological functioning [12,14-17]. These needs are closely linked to the domain of personal recovery [19]. The similarity lies in the aim to apply interventions to increase mental well-being, rather than just

symptomatic recovery. In personal recovery, there are 5 recognizable components: connectedness, hope, identity, meaning, and empowerment, indicated by the CHIME acronym [19]. Bird et al [20] added 3 components (practical support, issues around diagnosis and medication, and skepticism surrounding recovery) to CHIME to address the needs of those who are in an early stage of recovery. Several interventions have been developed based on the CHIME framework, focusing on increasing hope, strengths, connectedness, and empowerment [21,22]. A potential treatment approach to improve personal recovery and address the unmet needs of patients with BD is positive psychology (PP) [14,23-25].

PP is a relatively new field in psychology that focuses on improving positive feelings, behaviors, and cognitions [26-28]. Some important evidence-based types of PP interventions (PPIs) are *savoring* [29-32], *practicing kindness* [33], *experiencing and expressing gratitude* [34-38], *creating meaning and goal setting* [37,39-43], *positive relations* [23,37,44,45], and *using personal strengths* [36]. Meta-analyses have found that PPIs have small to moderately significant effects on well-being and distress in general populations [28,46-48] and clinical populations [49,50]. Some small studies have also shown promising effects of PPIs on the mental health of patients with BD [51-53]. Recently, a fully powered trial was conducted to evaluate the impact of a positive psychotherapy group treatment on the mental well-being of patients with BD in comparison with treatment as usual [54,55], reporting promising medium to large between-group effects, with sustained effect after 6 months [55]. Applying PPIs in mental health care fits well in a recently developed model of sustainable mental health [56]. In this model, both mental illness and mental well-being are proposed as vital outcomes in psychiatry. A core aim of treatment is to promote the patients' ability to adapt. This ability is hindered by barriers such as dysfunctional biological and psychological processes and enhanced by resources such as positive emotions, hope, meaning, and positive relationships. Bohlmeijer and Westerhof [57] argue that there is a need for balanced mental health care. PPIs primarily target the development of resources that support the patient with personal recovery and maintaining mental health.

Digital health interventions are increasingly common in mental health treatment. In PP, eHealth interventions are known as online PPIs (OPPIs). Studies in this area have shown that OPPIs can enhance well-being and reduce depressive symptoms [58]. Results from recent studies show a significant effect of OPPIs [59-61]. For those individuals who found that the interventions are relevant for their symptoms, OPPIs seem more acceptable [62]. It has also been found that the adjustability of a digital application improves its acceptability [62,63]. Patients with depressive symptoms seem to benefit more from OPPIs [64]. These findings suggest that digital applications can be a promising way to implement PPIs for patients with (bipolar) depression or those with low well-being levels.

Potential benefits of digital health interventions lie in the improved accessibility, flexibility in both standardization and personalization, interactivity, and consumer engagement [65]. Successful implementation of eHealth applications requires careful consideration of individual needs and cocreation with key stakeholders in both the design and implementation phases. However, adoption by users and professionals is not always easily achieved; professionals can be skeptical about the potential benefits and experience little support in using eHealth applications [66]. Implementing a web-based recovery treatment program for patients with severe mental illness revealed that they were not easily engaged [67]. However, these challenges should not prevent a push forward for health care technology changes [68]. Therefore, development and implementation of technical innovations require thorough communication and coordination between health care professionals and patients. In developing eHealth applications, user involvement is essential for technology adoption and use, increased user satisfaction, trust, and usability and is needed for successful implementation [69]. To achieve engagement, commitment, confidence, and a more positive attitude toward new eHealth interventions from potential users, van Gemert-Pijnen et al [70] developed a holistic approach for designing and implementing eHealth applications.

In summary, the overall burden of BD is vast and significantly impairs patients' quality of life. There are important unmet needs for patients with BD, which are mainly related to personal recovery. PP is a promising treatment approach to improve personal recovery. Implementing PPIs in digital interventions is potentially empowering and cost-effective. However, systematic user involvement in digital health interventions is vital. This study aimed to gain insight into the opinions of patients with BD and health care professionals about (web-based) PPIs for BD and to develop and pilot-test an app containing PPIs specifically designed for patients with BD.

The following stages were addressed in the study:

1. Assessment of opinions in Focus group meetings (contextual inquiry)
2. Assessment of preferences and requirements (value specification [VS] and design)
3. Assessment of use and satisfaction with positive psychology app and interventions (operationalization)

## Methods

### Design

The study focused on gathering information from patients and professionals about the use of PP to develop an app containing PPIs for patients with BD. A qualitative design was used to identify and characterize opinions and needs. The target group comprised patients with BD and professionals (end users). To enable a broad perspective on the potential benefits of web-based PPIs for patients with BD, focus group discussions (FGDs), paper prototyping (PPT), and pilot test (PT) were used. The method was modeled on the principles of the holistic development approach described in the Center for eHealth and

Disease Management (CeHRes) road map, for example, the participatory process and continuous evaluation cycles [70]. Our study covered the first 3 steps of the model: contextual inquiry, VS, and design. The CeHRes aims to facilitate continuing process of evaluation and participation of all stakeholders.

### Ethics Approval

Ethics approval was obtained from the University of Twente (18067) and the scientific board of the Dimence Mental Health Institute, where the study was conducted. All participants signed an informed consent form, in which they also agreed to be audio recorded during the focus groups (FGs).

### Participants

Participants were patients with BD I or II in a euthymic episode and professionals treating patients with BD who were in possession of a smartphone and were willing to travel to attend the FGs. They participated in all 3 steps of the development process. When recruiting our sample, we aimed for maximum variation among the participants of the FG. Therefore, we used several recruitment strategies. Patients were recruited from an outpatient clinic and the national advocacy group (Plusminus) to receive input from different regions of the Netherlands. In using this method of recruiting, we tried to avoid selection bias. All the patients received treatment as usual for BD. A call for participation was posted in the advocacy group's magazine. The researchers who participated in the FGs (BG and SK) were not involved in the participants' treatments. Professionals in the FG were involved in various disciplines (psychiatrists, psychiatric nurses, and psychologists) and recruited from the Dutch Foundation for Bipolar Disorders (Kenniscentrum Bipolaire Stoornissen), a chapter of the International Society for Bipolar Disorders.

Participants contributed to all study phases (FG, PPT, and PT); however, not all participants attended every FG. None of the participants in the FG had professional relationships, which means that none of the patients were treated by the professionals or otherwise involved with each other.

### Procedure and Materials

#### Overview

In total, three 2-hour FGDs were conducted in September 2018 and October 2018. At the beginning of the FG, a clear statement was made about confidentiality to ensure that the participants can independently provide their opinion. The sessions were semistructured, with the use of a topic list. In the first FGD, we explored experiences with PP. In the second FGD, we gathered the requirements for a potential PP app and the needs that will be covered by a PPI app. In the third FGD, the final requirements for the PPI app for BD were established. Each FG was briefly introduced with a PowerPoint (Microsoft Corp) presentation to reveal the purpose of that discussion. The participants were informed about the topics before the FG. [Table 1](#) provides an overview of the techniques used during the FGs.

**Table 1.** Overview of techniques used during the focus groups.

FGD <sup>a</sup>	Brown paper exercise	Paper prototyping	Rapid prototyping test	Valuation PP <sup>b</sup> exercises
1	✓			
2	✓	✓		✓
3			✓	

<sup>a</sup>FGD: focus group discussion.

<sup>b</sup>PP: positive psychology.

### **Assessment of Opinions in FMGs (Contextual Inquiry)**

After the brief introduction to the first FG, participants were asked to write down the experiences with PP, needs that PP could address, and requirements of a PPI app on post-it memos collected on flip charts (brown paper exercise). These statements were the basic assumptions used to initiate the discussion about the various opinions. The agreements and disagreements were summarized, and in the next FG, the members were asked again to mention a preference. This process was repeated until consensus was reached. The discussions were recorded and transcribed, followed by a consensus check. After each session, a report was made and sent to the participants for validation and comments to increase objectivity (member checking). Concepts of a *consensus document* were discussed in the second and third meetings and established after the last session.

### **Assessment of Preferences and Requirements (VS and Design)**

We also used the PPT method in FG 2 to test the preferences. On the basis of the preferences mentioned in the FG, we created a PP exercise (being grateful). The participants rated screenshots of the exercise in terms of content, wording, and design. Subsequently, we asked the participants their opinions and valuations about different PPIs to establish which PPIs seemed suitable for patients with BD. The 6 categories of PPI's (positive emotions, resilience, positive relations, optimism and hope, self-compassion, and strengths) were explained and practiced with 1 exercise per category. Then, the participants scored each type in a positive or negative appreciation. We conducted a rapid prototyping (RPT) test to establish the participants' opinions about a possible web-based PPI. On the basis of the first FG input, we built an exercise in 'The Incredible Intervention Machine,' an app specially designed to perform prototyping and pilot-testing of newly developed apps in a research setting [71].

On the basis of the results of the FGs, an app with PPIs (Well-being Bipolar Disorder) was designed. Then, this app was evaluated in a PT.

### **Assessment of Use and Satisfaction With PP App and Interventions (Operationalization)**

After the development process, we tested our app to evaluate whether the results of the previous steps of the development process had been implemented satisfactorily. Then, we maximized user involvement. In the PT, we tested the intervention app containing 7 PPI exercises in 1 week. The participants were asked to perform 1 exercise daily. After they completed 1 exercise, the following exercise appeared the next day. We collected data on the exercises separately through ranking after every exercise and the possibility to provide remarks about the exercise. We also collected data on the use of the app (preferences in settings, frequency of use, and number of completed exercises). After completing all the exercises, a final evaluation within the app was conducted to establish experiences about the intervention.

### **Data Analysis**

ATLAS.ti7 was used for the analysis of the data from the FG. The Colaizzi method, as described by Shosha [72], was used to process the data. The FG recordings were transcribed verbatim for the analysis in 3 phases: open coding, axial coding, and selective coding. Researcher triangulation was used to increase the objectivity of data analysis. The app's quantitative data (rating of the exercises) were collected, and average scores for each exercise and all the 7 exercises were calculated using SSPS. In addition, we also calculated the average scores of the differences between professionals and patients. The qualitative data (open answers in the evaluation module) were collected and analyzed through inductive coding.

## **Results**

### **Participants**

For the study, 17 participants (n=11, 65% patients and n=6, 35% professionals) were recruited. In total, 24% (4/17) of the participants withdrew (3/4, 75% patients and 1/4, 25% professional) before the start of the study owing to personal reasons or because they were not available at the time of the FGs. For the last phase (ie, PT of the app), 6 participants were added to broaden the input with opinions of participants who did not participate in the FGs. Demographics are shown in [Table 2](#). [Table 3](#) shows the participation in the FG, RPT, and PT.

**Table 2.** Demographics of participants of the focus groups.

Characteristics	Total (N=13), n (%)	Patients (n=8), n (%)	Professionals, (n=5), n (%)
<b>Age (years)</b>			
16-24	1 (8)	1 (13)	0 (0)
25-40	4 (31)	3 (38)	1 (20)
41-55	5 (38)	2 (25)	3 (60)
56-70	3 (23)	2 (25)	1 (20)
Sex (female)	7 (54)	4 (50)	3 (60)
<b>Marital status</b>			
Single	2 (15)	2 (25)	0 (0)
In relationship	6 (46)	4 (50)	2 (40)
In relationship and has children	5 (38)	2 (25)	3 (60)
<b>Education</b>			
Primary school	0 (0)	0 (0)	0 (0)
High school	3 (23)	3 (38)	0 (0)
Higher professional education	8 (62)	5 (63)	3 (60)
University	2 (15)	0 (0)	2 (40)

**Table 3.** Participation in the different stages of the study (N=19).

Part of the study	Patients (n=10), n (%)	Professionals (n=9), n (%)	Total (n=19), n (%)
FG <sup>a</sup> 1	4 (21)	2 (11)	6 (32)
FG 2	4 (21)	3 (16)	7 (37)
FG 3	4 (21)	4 (21)	8 (42)
PPT <sup>b</sup>	4 (21)	3 (16)	7 (37)
RPT <sup>c</sup>	4 (21)	4 (21)	8 (42)
PT <sup>d</sup>	10 (53)	9 (47)	19 (100)

<sup>a</sup>FG: focus group.

<sup>b</sup>PPT: paper prototyping.

<sup>c</sup>RPT: rapid prototyping.

<sup>d</sup>PT: pilot test.

## Focus Group Discussions

The results of all FGs are presented based on the different stages of the CeHRes road map. The results of the contextual inquiry and VS are summarized to provide a good overview of the FG results.

### Assessment of Opinions in FMGs (Contextual Inquiry)

In the FG, first, we discussed the participants' level of experience with PP. Second, we created an inventory of expected advantages and disadvantages when PPI is applied for BD. The results are summarized in [Table 4](#).

Among the FG members, there was little experience with PP, as shown in [Table 4](#). The participants did not have experiences with specific evidence-based PPIs. The experiences can be found in adjoining therapeutic areas (eg, mindfulness) or more personal PP solutions (eg, recognizing positive moments or thinking about possibilities rather than about problems). The FG members

mentioned possible advantages: focusing on small steps, making positive pictures or movies, monitoring positive feelings, and giving themselves a positive message. Unmet needs such as hope, acceptance, and increasing self-confidence appear to be the most promising ones that PPIs may address in BD. The participants also mentioned various potential disadvantages of PPIs in 2 categories: illness-related and personal factors. The expectation of disadvantages regarding their illness is seen in both mania and depressive episodes. The participants did not expect benefits from PPIs in severe depressive episodes, or even the inability to perform the exercises during severe depression, leading to disappointment rather than satisfaction. They also foresaw further mood dysregulation toward manic stages when they are already hypomanic owing to exercises that stimulate positive emotions, feelings, or happiness. On a personal level, the participants fear forced positive statements that are not consistent with their self-esteem and the risk that PPIs can push the goal setting level very far (perfectionism). For both

mentioned domains, they fear a possible counterproductive outcome when PPIs are applied.

**Table 4.** Contextual inquiry, experiences, and expectations.

Experiences	Number of times initially mentioned in brown paper exercise <sup>a</sup>	Expectations (advantage)	Unmet needs, as described in the literature [12,14-17]	Expectation's disadvantage	Label
Make mantras for yourself	1×	Avoid stigmatization	Social isolation and acceptance	Forced positive statements	Personal level
Exercising mindfulness	2×	Helpful in "gloomy" periods	Hope	Not beneficial when severely depressed	Illness-related factor
Caring for others	1×	Monitoring of positive feelings	Grief counseling and acceptance	Possibility of high goal setting	Personal level
Writing a "stoic journal" daily	1×	Focusing on small steps (near future)	Hope	Fear that positive feeling can lead to hypomania or mania	Illness-related factor
Knowing through reading about PP <sup>b</sup>	1×	Express gratitude	Expressing feelings	— <sup>c</sup>	—
Recognize positive moments	1×	Positive messages to yourself	Increasing self-confidence and hope	—	—
User involvement as a positive activity	1×	Create positive daily pictures or movies	Increasing self-confidence and hope	—	—
Thinking in possibilities	2×	—	—	—	—

<sup>a</sup>Participants wrote their thoughts and opinions on memo blocks sheets before the discussion started.

<sup>b</sup>PP: positive psychology.

<sup>c</sup>Not available.

### Assessment of Preferences and Requirements (VS)

The aim of the FGs was to address opinions about the use and requirements of an app. First, the participants were asked about their opinions on the different categories of PPIs (Table 5). The FG participants were unanimous in the appreciation of the exercises in positive emotions and resilience categories and found positive relations and strengths to be applicable for patients with BD. In the categories of self-compassion, hope, and optimism, the FG participants were less convinced; there was fear that they could not fulfill high expectations of themselves or the app demands. Second, a recurring topic in the discussion was personalization; individuals can have different preferences for exercises. Therefore, all types of exercises should be available within the app. In addition, the individual user can choose certain types of exercises. Third, some participants found it important that the choice was made in alignment with the professional caregiver. Fourth, we concretized the app's design on 2 levels: the app's use (VS) and its feel and look (design).

The most prominent subtopics in the use of the app were the following: when the app is used and under what conditions. Furthermore, the FG members made suggestions for more advanced use. These results are summarized in Table 6. The participants were unanimous that the best way to use OPPIs is during euthymic or mild depressive episodes. Participants were divided over the use of an OPPI in episodes of mania or hypomania. The FG members expected that use in the early stages of hypomania can provoke positive feelings and lead to a more severe manic state. However, when users are in a full

manic state, they did not expect any temptation to use the app in a full manic episode because, when manic, they are quickly distracted, and the exercises require time and tranquility.

The participants also indicated that the app should be adjustable; however, most FG members think that they want to use it daily, during periods when it is beneficial to them. The duration of the practices should be between 5 and 10 minutes, so that it fits into daily routines. The participants also mentioned that allowing sufficient time and having a quiet place is important for successfully using the app. In addition, the FG members suggested that, after performing a complete set of practices, the user can choose which practices are suitable and adjust the app to those preferences. In the discussion about the potential advanced possibilities, the FG members found it helpful to connect PPIs with the Life Chart Method (LCM) [73]. A PPI should occur when an advanced set value is reached while monitoring symptoms such as mild depression (within the LCM). Incorporating PPIs in the early relapse prevention plan was also suggested, as users can apply this intervention to handle the (starting) mood episode. In the ideal world, the FG members want to be automatically directed to the PPI when they reach pre-established levels in the LCM.

During FG 2 and 3, participants were asked about their views on the requirements for the design of the app. The remarks obtained can be separated into the following subcategories: personalization, look and feel, text, vision and sound, and preferred options. Table 7 summarizes the results.

The participants considered personalization to be a vital aspect. The FG members wanted to have a wide range of adjustable

options as long it does not affect the app’s clarity. The main topics in personalization were the frequency of use, choice between reading or listening, notifications, writing space in the exercises, and possibility to select the exercises. In addition, the proper use of wording was considered necessary by the participants. The text in the app should be concrete, clear, and short. Within the suggested exercises, the choice of words was sometimes perceived as compelling. Examples to illustrate the exercises were seen as helpful and supportive. The FG members had a preference for appealing messages in both exercises and notifications. They preferred a design that is quiet and attractive. The use of pictures can have a calming effect. We also discussed the option of obtaining feedback after completing an exercise. Approximately all participants (7/8, 88%) liked some type of feedback. They also expressed the need for external motivation to continue practicing (apart from the notifications). Significant others can be part of this motivational aspect. Valuing the exercises was seen as a good instrument to decide which exercises were preferable.

As part of the VS and design, we conducted an RPT test. Owing to technical problems in registration, 25% (2/8) of the participants did not succeed. The feedback obtained from the participants was divided into 4 categories: experienced effects, facilitating factors, impediments, and suggestions. Regarding experienced effects, 75% (6/8) of the participants found exercises to be beneficial and experienced more positive emotions than before. The experienced positive emotions did not qualify as ‘threatening’ in terms of risk for a (hypo)manic episode. One participant did not experience any difference, and another participant found it difficult to perform the exercise owing to personal circumstances. Regarding facilitating factors, the participants appreciated the layout with different pictures, videos, and music and the calm design. In addition, participants emphasized the possibility of adding the option to personalize the look and feel of the app. In total, 38% (3/8) of the participants experienced impediments while testing the app. Overall, 25% (2/8) of the participants mentioned that they were very severely depressed to perform the exercise.

**Table 5.** Valuation positive psychology interventions for use in bipolar disorder.

Theme	Positive remarks	Negative remarks	Appreciate (n=6), n (%)	Not appreciate (n=6), n (%)
Positive emotions	<ul style="list-style-type: none"> <li>“Creates freedom to concentrate on your positive emotions.”</li> </ul>	<ul style="list-style-type: none"> <li>“It’s difficult to allow yourself to do what you want to do.”</li> </ul>	6 (100)	0 (0)
Resilience	<ul style="list-style-type: none"> <li>“Gives energy.”</li> <li>“Achieve relaxation.”</li> <li>“Seeking solutions that fit me rather than always must to...”</li> </ul>	<ul style="list-style-type: none"> <li>“Right wording is essential.”</li> </ul>	6 (100)	0 (0)
Positive relations	<ul style="list-style-type: none"> <li>“Focus on connecting with other people.”</li> <li>“The positive contact with others inspires me.”</li> </ul>	<ul style="list-style-type: none"> <li>“Contacts have to be trusted before sharing feelings.”</li> </ul>	5 (83)	0 (0)
Strengths	<ul style="list-style-type: none"> <li>“No remarks were made.”</li> </ul>	<ul style="list-style-type: none"> <li>“I don’t give myself time for that either.”</li> <li>“Right wording is essential.”</li> </ul>	5 (83)	0 (0)
Self-compassion	<ul style="list-style-type: none"> <li>“Allow yourself to comfort yourself.”</li> </ul>	<ul style="list-style-type: none"> <li>“It’s problematic to allow yourself to give some consolation.”</li> <li>“It mustn’t be compulsory.”</li> <li>“Difficult to perform the exercises.”</li> </ul>	4 (67)	2 (33)
Optimism and hope	<ul style="list-style-type: none"> <li>“Hope is important; perhaps the exercise doesn’t fit.”</li> </ul>	<ul style="list-style-type: none"> <li>“Doesn’t fit people with perfectional traits.”</li> <li>“High expectations that maybe can’t be satisfied.”</li> <li>“It seems to be a bit ‘Trump-like.’”</li> <li>“Too ambitious.”</li> </ul>	3 (50)	2 (33)



**Table 6.** Design and overview of the use of web-based positive psychology interventions for bipolar disorder.

Subcategories in using and remarks made by FG <sup>a</sup> members	Agreed by all FG members
<b>When to use the app</b>	
“In gloomy periods but not in severe depressive episodes.” <sup>b</sup>	Yes
“In mild hypomanic episodes.” <sup>c</sup>	Yes
“In euthymic episodes.”	Yes
“Suggest the user exercises on fixed times.”	No
“Adjustable frequency of the exercises.”	Yes
<b>How to use the app</b>	
“Work through all exercises; eg, 1 exercise every day for 6 weeks and then integrate it into the Life/Chart.”	Partly
“In the beginning, the user goes through a module with practices of the themes; positive emotions, positive relationships and resilience.” and “the themes hope, and optimism, strengths and self-compassion are offered as an option.”	Yes
“Suggest clearly to do the exercises in a safe environment.”	No
“Set up realistic goal setting.”	Yes
“The duration is between 5-10 minutes per exercise.”	Yes
<b>Advance possibilities</b>	
“Connection with the LCM <sup>d</sup> .”	Yes
“Connection with the relapse prevention plan or other recovery plans.”	Yes
“Going through the different exercises with the practitioner to choose a set of exercises.”	No

<sup>a</sup>FG: focus group.

<sup>b</sup>Compared with Life Chart Method—mild or moderate depression.

<sup>c</sup>Compared with Life Chart Method—mild hypomanic episodes.

<sup>d</sup>LCM: Life-Chart Method.

**Table 7.** Design and overview of the feel and look requirements.

Subcategories and remarks made by FG <sup>a</sup> members	Agreed by all FG members
<b>Personalization</b>	
“The choice between reading or listening to the exercise.”	Yes
“Pleasant voice; voices can be chosen.”	Yes
“Space to type keywords within the exercises.”	No
“Ability to select which exercise you want to do.”	Yes
“The degree of customization must be large, but the app must remain clear to promote easy use.”	Yes
“The notifications must be flexible, with the option of carrying out the exercise later.”	Yes
“Personalization, not only in exercises but also in the used pictures, videos and music fragments.”	Yes
<b>Look and feel—text</b>	
“There must be a choice between spoken or written exercises.”	Yes
“Use not too many words; make clear short exercises.”	Yes
“The text should be inviting with a smooth choice of words but not too clever and easy to read.”	Yes
“Working with examples in the exercises.”	No
“Limited the amount of text per screen.”	Yes
<b>Look and feel—vision</b>	
“Quiet design, with a nice layout and images.”	Yes
“Use appropriate images for the exercises.”	Yes
“Offer the possibility to add pictures yourself.”	Yes
“Be visually appealing; photos/graphics.”	Yes
“Use animations for the explanation in the exercises.”	No
<b>Look and feel—sound</b>	
“Pleasant voice; voices can be chosen.”	Yes
“There must be a choice between spoken or written exercises.”	Yes
“Possibility to add music.”	No
<b>Preferred options</b>	
“Feedback, compliments after every completed exercise.”	Partly <sup>b</sup>
“Selection menu for the exercises.”	Yes
“Add your own exercises (in a simple layout or only as a reminder).”	No
“Being able to give a score yourself and make this visible in a graph.”	No
“To be able to share the results of the exercises with others.”	Yes
“Receive an anonymous response from others—as a tip or encouragement, which must be adjustable.”	No

<sup>a</sup>FG: focus group.

<sup>b</sup>A participant did not like the option to obtain feedback in the app.

### Assessment of Use and Satisfaction With PP App and Interventions (Design and Operationalization)—Exercises of the Intervention

On the basis of the consensus reached in the FG and results of the RPT, we developed the Wellbeing Bipolar Disorder app. The app contains 7 exercises in the 4 domains of PP preferred by FG members. The exercises based on the previous work of Bohlmeijer and Hulsbergen [74] are shown in Table 8. The app differs from other apps that provide PPIs, primarily owing to the selection of PPIs. Some exercises were withdrawn because

of the fear of compassion or had to be altered. Other exercises were withdrawn because they may provoke symptoms of mania. Moreover, the app had special design features, such as calm design, proper use of wording, and so on, which, according to the FG members, can lead to high compliance rates. Finally, the app can be integrated into the LCM.

We chose a 1-week period in which the participants were provided 1 exercise daily, with mood monitoring before and after exercise and a valuation of the exercise after completing it. The app was personalized with the following choices: when to use the app, notifications, types of notifications, written or

spoken video explanation, and guidance by a professional or expert by experience. The app's design was calm. All exercises had 1 picture throughout the practice (Figure 1).

The participants were asked if they had succeeded in the exercise; if they did not, they could start again. We wrote 2 different scripts (professional or expert by experience), and the videos were set in the background picture belonging to the exercise.

After the development process, we tested our app to evaluate whether the results of the previous steps of the development process have been implemented to the participants' satisfaction. The results of the PT are shown in Tables 9 and 10.

Of the 133 exercises (7×19 participants), 87 (65.4%) were completed. Patients completed more (101/133, 75.9%) exercises than professionals (72/133, 54.1%). In total, 11% (2/19) of the participants chose the video explanation. Of the 19 participants, 8 (42%) preferred the expert by experience and 8 (42%) preferred the guidance by a professional. Overall, the average rating of all exercises in total was 7.35 (scale 0-10, SD 0.525),

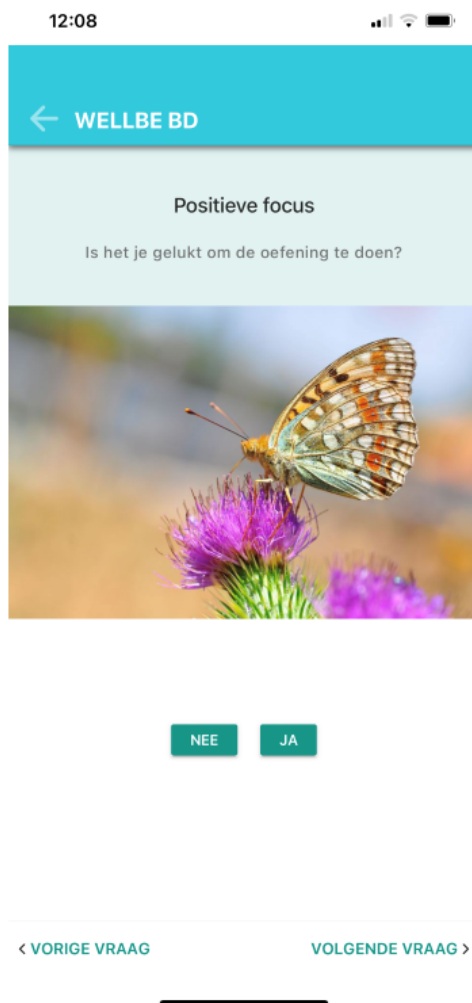
and the median was 7.5, with a slightly high rating among professionals (mean 7.7 vs 6.9; median 7.5 vs 7.25).

The evaluation of the individual exercises was between 7.5 (exercise: be strong and be stronger) and 6.6 (exercise; listening to good news). Notably, a participant rated all exercises relatively low (average 3.3), and owing to the small sample size, this influenced the total outcome. The other individual ratings were between 8.7 and 5.6.

The app's overall valuation was high; 91% (15/16) of the participants were positive about the app and wanted to use it for an extended period. However, according to 55% (9/16) of the participants, the frequency of the exercises seemed very high for an extended period. Besides positive comments, made in the app's evaluation, about the effect of the exercises, there were remarks for improvement of the app. Some comments referred to the intensity of the exercises; a new exercise every day is not doable for all participants. Another advised option is to read the exercises in the morning to accomplish it during the day.

**Table 8.** Domains and exercises in the Well-being Bipolar Disorder app.

Exercise number	Exercise	Domain of positive psychology
1	Experience positive moments again	Positive emotions
2	Active listening to good news	Positive relations
3	Three good things exercise	Resilience
4	Discover your strengths	Strengths
5	Positive focus	Positive emotions
6	Expressing gratitude	Positive relations
7	Being strong and becoming stronger	Resilience

**Figure 1.** Screenshot of the intervention.**Table 9.** Overview of the outcome of pilot test of the Well or being Bipolar Disorder app I.

Exercise number	Number accomplished <sup>a</sup>	Valuation by patients (scale 0-10) <sup>b</sup>	Valuation by professionals (scale 0-10) <sup>c</sup>	Valuation—total (scale 0-10), mean (SD) <sup>d</sup>
1	14	7.37	7.5	7.43 (0.650)
2	12	5.7	7.8	6.58 (1.054)
3	13	7.37	8.2	7.69 (0.418)
4	13	6.87	7	6.92 (0.065)
5	13	7.22	8	7.38 (0.411)
6	11	6.28	8	6.91 (0.870)
7	11	7.42	7.8	7.45 (0.211)

<sup>a</sup>Total number accomplished=87.

<sup>b</sup>Valuation by patients (total)=6.85.

<sup>c</sup>Valuation by professionals (total)=7.7.

<sup>d</sup>Valuation by all participants (total)=7.25.

**Table 10.** Overview of the outcome of pilot test of the Well-being Bipolar Disorder app II.

Participants	Percentages of accomplished exercises <sup>a</sup>	Valuation of the app (positive) <sup>b</sup> , %	More extended use of the app (positive) <sup>c</sup> , %	Frequency to high <sup>d</sup> , %
Patients	(75.7)	83	83	30
Professionals	(54)	100	100	80

<sup>a</sup>Number of exercises (total)=65.4%.

<sup>b</sup>Valuation of the app (positive; total)=91.5%.

<sup>c</sup>More extended use of the app (positive; total)=91.5%.

<sup>d</sup>Frequency to high (total)=55%, SD 0.525%.

## Discussion

### Principal Findings

This study investigated the opinions of care professionals and patients with BD regarding OPPI for BD.

The first aim was to rate the extent to which participants had experience with PPIs and expected them to be supportive in addressing unmet needs. We found that the participants did not have experience with specific evidenced PPIs; however, they mentioned experiences in adjoining therapeutic areas (eg, mindfulness) or more personal PP solutions. The FG members mentioned possible advantages: focusing on small steps, making positive pictures or movies, monitoring positive feelings, and giving yourself a positive message. These topics can be related to various unmet needs in treating BD, specifically, offering hope, increasing self-esteem, expressing feelings, promoting acceptance, and preventing social isolation. Regarding expectations about PPIs, participants expect that PPIs can accomplish some unmet needs in BD.

The second aim was to establish the preferences and requirements of the app with PPIs.

First, we determined which PP exercises patients and professionals prefer. Positive emotions, resilience, positive relations, and strengths were highly valued among the 6 categories of PP exercises. Some exercises were withdrawn due to the fear of compassion or had to be altered. Other exercises were withdrawn because they may provoke symptoms of mania. Moreover, the app had special design features, such as calm design, proper use of wording, and so on, which, according to the FG members, can lead to high compliance rates. Finally, the app can be integrated into the LCM. Therefore, we developed an app that differs from other apps that provide PPIs, primarily owing to the selection of PPIs.

The third aim of our study was to evaluate the use and satisfaction of the app. An interesting finding is that approximately all users (15/16, 91%) found it to be beneficial to perform the exercises and wanted to do it regularly. However, the frequency of the exercises seems to be very high. The valuation of the exercises was promising (7.35 on a scale of 1-10; median 7.5). Despite the small number of participants, we seemed to have found the proper exercises for our target group. Before releasing the app for clinical practice, further studies with adequate measurements (quantitative and qualitative) are necessary.

### Comparison With Previous Studies

Unmet needs can include topics such as support with loneliness, grief counseling, acceptance, social isolation, coping with others in social functioning [14,15,17], hope, expressing feelings, and increasing self-confidence in psychological functioning [12,14-17]. The positive expectations underscore the potential of PP for personal recovery and support the value of integrating PPIs into mental health care [57,58]. PPIs may help patients with resources that increase their ability to adapt and support them in personal recovery [24]. Our results are largely consistent with previous findings [19,75,76]. Mental well-being is recognized as an essential resource for the recovery from mental illness and in preventing relapse. Therefore, it is recommended to include mental well-being interventions (such as PPI) in the treatment [75,77].

Participants also raised some critical concerns about applying PPIs for BD. The FG members expressed concerns about fast and more severe changes in mood and energy when a used PPIs during manic or hypomanic episodes. Although “joy and amusement” are associated with increased manic severity, compassion—one of the key elements of PP—tends to decrease the symptoms of mania [78]. Positive emotions are not often mentioned as triggers that can provoke manic episodes. Lack of sleep is the most predominant factor in triggering manic episodes [79]. However, amorousness is the predominant factor among young adults, followed by stressful life events [80]. Periods of strong personal growth are also factors that induce manic symptoms (such as an extremely motivational workshop) [80]. Although this seems to be linked to PPIs, it is not satisfactory to conclude that PPIs can induce manic episodes. In a study among a large population of patients with BD (n=149), by applying a group PPI, the researchers found no increase in manic symptoms, thus confirming our findings [55]. In contrast, damping emotions lead to more severe depressive symptoms [81]. Therefore, it is recommended to inform potential participants properly before applying PPIs for BD. Interestingly, the PT results did not show dysregulation among the participants.

An unexpected finding was that the categories of self-compassion, optimism, and hope had low ratings. On the basis of the unmet needs, as described previously, we did not expect this outcome. This result may be explained by the fact that the FG members found the themes and exercises to be very ambitious. For example, the exercise, “the best possible self,” seems not to fit with the participants’ level of self-esteem. They had high expectations that they could not satisfy.

The “fear of compassion” can explain the fact that the FG members valued compassion relatively low. The fear of compassion is closely linked to self-criticism and depression [81]. Women with BD seem to be more self-critical than controls [82]. Nitzburg et al [83] suggest that the negative experiences in the course of BD can worsen the level of self-criticism and argue that in an early stage of the illness, interventions should also target decreasing self-criticism. However, owing to the small sample size, we cannot rule out that our findings are coincidental. Nevertheless, it is necessary to develop interventions targeting compassion and hope in BD to pay attention to the fear of compassion and transform exercises in a feasible manner for patients with BD.

We determined when to use the app. Some of the FG members expect the exercises to be a daily routine, for example, in addition to mood monitoring. Integration with a (digital) mood monitoring application and relapse prevention plan can provide a comprehensive tool in which the PPIs assist in preventing severe mood episodes [84]. The FG members foresaw the use mainly during euthymic or mild depressive episodes. The members of the FGs were unanimous in that they did not expect any beneficial effects from PPI in severe depressive episodes. In contrast, they expected the opposite effect: worsening of depressive feelings if the goals of the exercises are not achieved. Previous meta-analyses have demonstrated that PPIs seem to be more effective in populations that are nondepressed or mildly depressed [28,50] than in populations with major mood disorders [50]. This supported the FG members’ suggestion to apply PPI only in mild depressive episodes and euthymic episodes. Carr et al [85] published a meta-analysis showing that in patients with depression, PPIs had a small to moderate effect on depression in terms of symptom reduction. However, in that paper, the severity of the depression was unclear. The considerations mentioned previously can underline our finding that personalization is one of the main topics in designing a PPI app for BD. Lack of personalization is, among with depressive symptoms, an important barrier to adapting digital mental health interventions [86]. Recognizing and resolving these barriers in the development process can contribute to the adaptation of new interventions [86].

Finally, we reflected on the method. We recognized that cocreation and user involvement are important in developing an app to be accepted by end users [87]. Specific findings, such as the preferences for the exercise categories or the possible risks in applying PPIs, will not be revealed, as we developed the app without user involvement. The participation of end users is the main element in user-centered design, defined by Preece et al [88] as “an approach, which views knowledge about users and their involvement in the design process as a central concern.” The development process conformed to the CeHRes principles [70]. This method allowed us to systematically involve users (patients and professionals) in the study’s development process. This enabled us to modify the app’s design through all stages of the development process and guaranteed maximum involvement of all stakeholders. Previous studies using the same principle support our findings [88]. The development of new interventions benefits from user involvement in all stages to meet the target group’s needs [89].

## Limitations

Our study has some limitations. When recruiting professionals, there is a risk of selection bias owing to the voluntary recruitment from a highly specialized professional pool (the Dutch Foundation for Bipolar Disorders). It is conceivable that the professionals interested in web-based monitoring are stepping forward to participate in this study. However, this method was chosen to obtain professional input from different parts of the Netherlands to avoid inputs from only one region. The same applies to the consumers; they were partly recruited from an outpatient clinic and partly from the Dutch advocacy network, “Plusminus.” Although this may raise the question of whether the participants represent the target group, we assume that, by combining advocacy members and patients treated in an outpatient clinic, our study is sufficiently representative.

The use of FGs has some limitations, such as the possibility of “group effect,” in which patients tend to adapt to the group leading opinions. Therefore, it is difficult to separate a personal opinion from a group opinion [90]. We tried to avoid this bias by collecting individual data (post-it memos) before the discussion in the group. Furthermore, it is sometimes difficult to generalize the outcome of FGs [90]. We tried to avoid this bias as much as possible through nationwide recruitment.

Finally, this study is mainly a qualitative study with a relatively small number of participants; therefore, it remains uncertain whether the results are sufficiently generalizable.

Despite these limitations, we believe that we shed light on consumers’ and professionals’ thoughts and considerations about using PP apps for BD.

## Conclusions and Practical Implications

Despite recognizing the possible benefits of PPI in BD and that they may address unmet needs in BD, very little is known about the effect of applying PPI in the treatment of BD.

In this study, we realized the shared assumptions about the application of PPIs for BD. The consensus on the different topics regarding the use of PPI shows that both patients and professionals underline the beneficial possibility of applying PPIs for BD. The use during subsyndromal and mild depressive episodes seems to be the most fruitful period for patients with BD. We did not establish the risk of provoking mania or hypomania by performing PPIs, but we could not draw firm conclusions because of the small sample size.

With the systematic use of user involvement (patients and professionals) in all steps of the development process, we were able to create an app that can potentially fulfill some of the current unmet needs in the treatment of BS.

The use of PPI for BD is intriguing and can be usefully explored in further studies. We emphasize that more evaluation studies (quantitative and qualitative) that are focused on the effect of PPIs in the treatment of BD should be conducted. In addition, to establish the working mechanisms in BD, explorative, qualitative, designed studies are required to reveal whether PPIs can address unmet needs in BD.

## Acknowledgments

There is no financial disclosure; sponsors were not involved in the study. The study was funded by the Dimence Mental Health Institute and the University of Twente.

## Authors' Contributions

BG is the first author, SMK is the first supervisor, AWMMS is the reviewer, RWK is the senior reviewer, and ETB is the senior reviewer and supervisor.

## Conflicts of Interest

None declared.

## References

1. Goodwin F, Jamison K. Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression, Second Edition. New York: Oxford University Press; 2007.
2. Geddes J, Miklowitz D. Treatment of bipolar disorder. *Lancet* 2013 May;381(9878):1672-1682 [FREE Full text] [doi: [10.1016/s0140-6736\(13\)60857-0](https://doi.org/10.1016/s0140-6736(13)60857-0)]
3. de Graaf R, ten Have M, van Dorsselaer S. De psychische gezondheid van de Nederlandse bevolking. Utrecht: Trimbos-instituut; 2010.
4. Erten E, Alpman N, Özdemir A, Fıstıkcı N. [The impact of disease course and type of episodes in bipolar disorder on caregiver burden]. *Turk Psikiyatri Derg* 2014;25(2):114-123 [FREE Full text] [Medline: [24936758](https://pubmed.ncbi.nlm.nih.gov/24936758/)]
5. Ferrari A, Stockings E, Khoo J, Erskine H, Degenhardt L, Vos T, et al. The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. *Bipolar Disord* 2016 Aug 27;18(5):440-450 [FREE Full text] [doi: [10.1111/bdi.12423](https://doi.org/10.1111/bdi.12423)]
6. Vieta E, Langosch J, Figueira M, Souery D, Blasco-Colmenares E, Medina E, et al. Clinical management and burden of bipolar disorder: results from a multinational longitudinal study (WAVE-bd). *Int J Neuropsychopharmacol* 2013 Sep;16(8):1719-1732 [FREE Full text] [doi: [10.1017/S1461145713000278](https://doi.org/10.1017/S1461145713000278)] [Medline: [23663490](https://pubmed.ncbi.nlm.nih.gov/23663490/)]
7. Kasper S. Living with bipolar disorder. *Expert Rev Neurotherapeutics* 2014 Jan 10;4(sup2):S9-15 [FREE Full text] [doi: [10.1586/14737175.4.6.s9](https://doi.org/10.1586/14737175.4.6.s9)]
8. Kupka R, Duffy A, Scott J, Almeida J, Balanzá-Martínez V, Birmaher B, et al. Consensus on nomenclature for clinical staging models in bipolar disorder: a narrative review from the International Society for Bipolar Disorders (ISBD) Staging Task Force. *Bipolar Disord* 2021 Nov;23(7):659-678 [FREE Full text] [doi: [10.1111/bdi.13105](https://doi.org/10.1111/bdi.13105)] [Medline: [34174130](https://pubmed.ncbi.nlm.nih.gov/34174130/)]
9. Berk M, Conus P, Lucas N, Hallam K, Malhi G, Dodd S, et al. Setting the stage: from prodrome to treatment resistance in bipolar disorder. *Bipolar Disord* 2007 Nov;9(7):671-678 [FREE Full text] [doi: [10.1111/j.1399-5618.2007.00484.x](https://doi.org/10.1111/j.1399-5618.2007.00484.x)] [Medline: [17988356](https://pubmed.ncbi.nlm.nih.gov/17988356/)]
10. Gershon A, Eidelman P. Inter-episode affective intensity and instability: predictors of depression and functional impairment in bipolar disorder. *J Behav Ther Exp Psychiatry* 2015 Mar;46:14-18 [FREE Full text] [doi: [10.1016/j.jbtep.2014.07.005](https://doi.org/10.1016/j.jbtep.2014.07.005)] [Medline: [25164093](https://pubmed.ncbi.nlm.nih.gov/25164093/)]
11. Fagiolini A, Kupfer D, Masalehdan A, Scott J, Houck P, Frank E. Functional impairment in the remission phase of bipolar disorder. *Bipolar Disord* 2005 Jun;7(3):281-285 [FREE Full text] [doi: [10.1111/j.1399-5618.2005.00207.x](https://doi.org/10.1111/j.1399-5618.2005.00207.x)] [Medline: [15898966](https://pubmed.ncbi.nlm.nih.gov/15898966/)]
12. Bauer M, Andreassen O, Geddes J, Vedel Kessing L, Lewitzka U, Schulze T, et al. Areas of uncertainties and unmet needs in bipolar disorders: clinical and research perspectives. *Lancet Psychiatry* 2018 Nov;5(11):930-939 [FREE Full text] [doi: [10.1016/s2215-0366\(18\)30253-0](https://doi.org/10.1016/s2215-0366(18)30253-0)]
13. Chengappa KR, Williams P. Barriers to the effective management of bipolar disorder: a survey of psychiatrists based in the UK and USA. *Bipolar Disord* 2005;7 Suppl 1:38-42 [FREE Full text] [doi: [10.1111/j.1399-5618.2005.00193.x](https://doi.org/10.1111/j.1399-5618.2005.00193.x)] [Medline: [15762867](https://pubmed.ncbi.nlm.nih.gov/15762867/)]
14. Fortuna K, Ferron J, Pratt S, Muralidharan A, Aschbrenner K, Williams A, et al. Unmet needs of people with serious mental illness: perspectives from certified peer specialists. *Psychiatr Q* 2019 Sep;90(3):579-586 [FREE Full text] [doi: [10.1007/s11126-019-09647-y](https://doi.org/10.1007/s11126-019-09647-y)] [Medline: [31154551](https://pubmed.ncbi.nlm.nih.gov/31154551/)]
15. Goossens P, Knoppert-van der Klein EA, Kroon H, van Achterberg T. Self-reported care needs of outpatients with a bipolar disorder in the Netherlands. *J Psychiatr Ment Health Nurs* 2007 Sep;14(6):549-557 [FREE Full text] [doi: [10.1111/j.1365-2850.2007.01127.x](https://doi.org/10.1111/j.1365-2850.2007.01127.x)] [Medline: [17718727](https://pubmed.ncbi.nlm.nih.gov/17718727/)]
16. Prasko J, Hajda M, Latalova K, Hruba R, Ociskova M, Holubova M, et al. Unmet needs of bipolar disorder patients. *Neuropsychiatric Disease Treatment* 2016 Jun;Volume 12:1561-1570 [FREE Full text] [doi: [10.2147/ndt.s105728](https://doi.org/10.2147/ndt.s105728)]

17. Maassen E, Regeer B, Regeer E, Bunders J, Kupka R. The challenges of living with bipolar disorder: a qualitative study of the implications for health care and research. *Int J Bipolar Disord* 2018 Nov 06;6(1):23 [FREE Full text] [doi: [10.1186/s40345-018-0131-y](https://doi.org/10.1186/s40345-018-0131-y)] [Medline: [30397833](https://pubmed.ncbi.nlm.nih.gov/30397833/)]
18. Asadi-Lari M, Tamburini M, Gray D. Patients' needs, satisfaction, and health related quality of life: towards a comprehensive model. *Health Qual Life Outcomes* 2004 Jun 29;2:32 [FREE Full text] [doi: [10.1186/1477-7525-2-32](https://doi.org/10.1186/1477-7525-2-32)] [Medline: [15225377](https://pubmed.ncbi.nlm.nih.gov/15225377/)]
19. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *Br J Psychiatry* 2011 Dec;199(6):445-452 [FREE Full text] [doi: [10.1192/bjp.bp.110.083733](https://doi.org/10.1192/bjp.bp.110.083733)] [Medline: [22130746](https://pubmed.ncbi.nlm.nih.gov/22130746/)]
20. Bird V, Leamy M, Tew J, Le Boutillier C, Williams J, Slade M. Fit for purpose? Validation of a conceptual framework for personal recovery with current mental health consumers. *Aust N Z J Psychiatry* 2014 Jul;48(7):644-653 [FREE Full text] [doi: [10.1177/0004867413520046](https://doi.org/10.1177/0004867413520046)] [Medline: [24413806](https://pubmed.ncbi.nlm.nih.gov/24413806/)]
21. Schrank B, Bird V, Rudnick A, Slade M. Determinants, self-management strategies and interventions for hope in people with mental disorders: systematic search and narrative review. *Soc Sci Med* 2012 Feb;74(4):554-564 [FREE Full text] [doi: [10.1016/j.socscimed.2011.11.008](https://doi.org/10.1016/j.socscimed.2011.11.008)] [Medline: [22240450](https://pubmed.ncbi.nlm.nih.gov/22240450/)]
22. Tew J, Ramon S, Slade M, Bird V, Melton J, Le Boutillier C. Social factors and recovery from mental health difficulties: a review of the evidence. *Brit J Social Work* 2011 Jun 15;42(3):443-460 [FREE Full text] [doi: [10.1093/bjsw/bcr076](https://doi.org/10.1093/bjsw/bcr076)]
23. Kraiss J, Ten Klooster PM, Frye E, Kupka R, Bohlmeijer E. Exploring factors associated with personal recovery in bipolar disorder. *Psychol Psychother* 2021 Sep;94(3):667-685 [FREE Full text] [doi: [10.1111/papt.12339](https://doi.org/10.1111/papt.12339)] [Medline: [33742536](https://pubmed.ncbi.nlm.nih.gov/33742536/)]
24. Slade M. Mental illness and well-being: the central importance of positive psychology and recovery approaches. *BMC Health Serv Res* 2010 Jan 26;10:26 [FREE Full text] [doi: [10.1186/1472-6963-10-26](https://doi.org/10.1186/1472-6963-10-26)] [Medline: [20102609](https://pubmed.ncbi.nlm.nih.gov/20102609/)]
25. Slade M, Amering M, Farkas M, Hamilton B, O'Hagan M, Panther G, et al. Uses and abuses of recovery: implementing recovery-oriented practices in mental health systems. *World Psychiatry* 2014 Feb;13(1):12-20 [FREE Full text] [doi: [10.1002/wps.20084](https://doi.org/10.1002/wps.20084)] [Medline: [24497237](https://pubmed.ncbi.nlm.nih.gov/24497237/)]
26. Parks A, Biswas-Diener R. Positive interventions: past, present, and future. In: *Mindfulness, Acceptance, and Positive Psychology: the Seven Foundations of Well-being*. Oakland, California, United States: New Harbinger Publications, Inc; 2013.
27. Seligman M, Csikszentmihalyi M. Positive psychology: an introduction. *Am Psychol* 2000;55(1):5-14 [FREE Full text] [doi: [10.1037/0003-066x.55.1.5](https://doi.org/10.1037/0003-066x.55.1.5)]
28. Sin N, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. *J Clin Psychol* 2009 May;65(5):467-487 [FREE Full text] [doi: [10.1002/jclp.20593](https://doi.org/10.1002/jclp.20593)] [Medline: [19301241](https://pubmed.ncbi.nlm.nih.gov/19301241/)]
29. Bryant F. Savoring Beliefs Inventory (SBI): a scale for measuring beliefs about savouring. *J Mental Health* 2009 Jul 06;12(2):175-196 [FREE Full text] [doi: [10.1080/0963823031000103489](https://doi.org/10.1080/0963823031000103489)]
30. Bryant F, Smart C, King S. Using the past to enhance the present: boosting happiness through positive reminiscence. *J Happiness Stud* 2005 Sep;6(3):227-260 [FREE Full text] [doi: [10.1007/s10902-005-3889-4](https://doi.org/10.1007/s10902-005-3889-4)]
31. Bryant F, Veroff J. Enhancing savoring. In: *Savoring*. London, England, United Kingdom: Psychology Press; 2007.
32. Schueller S. Preferences for positive psychology exercises. *J Positive Psychol* 2010 May;5(3):192-203 [FREE Full text] [doi: [10.1080/17439761003790948](https://doi.org/10.1080/17439761003790948)]
33. Otake K, Shimai S, Tanaka-Matsumi J, Otsui K, Fredrickson B. Happy people become happier through kindness: a counting kindnesses intervention. *J Happiness Stud* 2006 Sep;7(3):361-375 [FREE Full text] [doi: [10.1007/s10902-005-3650-z](https://doi.org/10.1007/s10902-005-3650-z)] [Medline: [17356687](https://pubmed.ncbi.nlm.nih.gov/17356687/)]
34. Boehm J, Lyubomirsky S, Sheldon K. A longitudinal experimental study comparing the effectiveness of happiness-enhancing strategies in Anglo Americans and Asian Americans. *Cogn Emot* 2011 Nov;25(7):1263-1272 [FREE Full text] [doi: [10.1080/02699931.2010.541227](https://doi.org/10.1080/02699931.2010.541227)] [Medline: [21432648](https://pubmed.ncbi.nlm.nih.gov/21432648/)]
35. Lambert N, Clark M, Durtschi J, Fincham F, Graham S. Benefits of expressing gratitude: expressing gratitude to a partner changes one's view of the relationship. *Psychol Sci* 2010 Apr;21(4):574-580 [FREE Full text] [doi: [10.1177/0956797610364003](https://doi.org/10.1177/0956797610364003)] [Medline: [20424104](https://pubmed.ncbi.nlm.nih.gov/20424104/)]
36. Seligman M, Steen T, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. *Am Psychol* 2005;60(5):410-421 [FREE Full text] [doi: [10.1037/0003-066x.60.5.410](https://doi.org/10.1037/0003-066x.60.5.410)]
37. Sheldon K, Lyubomirsky S. How to increase and sustain positive emotion: the effects of expressing gratitude and visualizing best possible selves. *J Positive Psychol* 2006 Apr;1(2):73-82 [FREE Full text] [doi: [10.1080/17439760500510676](https://doi.org/10.1080/17439760500510676)]
38. Wood A, Froh J, Geraghty A. Gratitude and well-being: a review and theoretical integration. *Clin Psychol Rev* 2010 Nov;30(7):890-905 [FREE Full text] [doi: [10.1016/j.cpr.2010.03.005](https://doi.org/10.1016/j.cpr.2010.03.005)] [Medline: [20451313](https://pubmed.ncbi.nlm.nih.gov/20451313/)]
39. Green L, Oades LG, Grant AM. Cognitive-behavioral, solution-focused life coaching: enhancing goal striving, well-being, and hope. *J Positive Psychol* 2006 Jul;1(3):142-149 [FREE Full text] [doi: [10.1080/17439760600619849](https://doi.org/10.1080/17439760600619849)]
40. King L. The health benefits of writing about life goals. *Pers Soc Psychol Bull* 2016 Jul 02;27(7):798-807 [FREE Full text] [doi: [10.1177/0146167201277003](https://doi.org/10.1177/0146167201277003)]
41. MacLeod A, Coates E, Hetherington J. Increasing well-being through teaching goal-setting and planning skills: results of a brief intervention. *J Happiness Stud* 2007 May 10;9(2):185-196 [FREE Full text] [doi: [10.1007/s10902-007-9057-2](https://doi.org/10.1007/s10902-007-9057-2)]



42. Sheldon K, Kasser T, Smith K, Share T. Personal goals and psychological growth: testing an intervention to enhance goal attainment and personality integration. *J Pers* 2002 Feb;70(1):5-31 [FREE Full text] [doi: [10.1111/1467-6494.00176](https://doi.org/10.1111/1467-6494.00176)] [Medline: [11908535](https://pubmed.ncbi.nlm.nih.gov/11908535/)]
43. Snyder C. TARGET ARTICLE: hope theory: rainbows in the mind. *Psychol Inquiry* 2002 Oct;13(4):249-275 [FREE Full text] [doi: [10.1207/s15327965pli1304\\_01](https://doi.org/10.1207/s15327965pli1304_01)]
44. Gable S, Reis HT, Impett EA, Asher E. What do you do when things go right? The intrapersonal and interpersonal benefits of sharing positive events. *J Pers Soc Psychol* 2004 Aug;87(2):228-245 [FREE Full text] [doi: [10.1037/0022-3514.87.2.228](https://doi.org/10.1037/0022-3514.87.2.228)] [Medline: [15301629](https://pubmed.ncbi.nlm.nih.gov/15301629/)]
45. Reivich K, Seligman M, McBride S. Master resilience training in the U.S. Army. *Am Psychol* 2011 Jan;66(1):25-34 [FREE Full text] [doi: [10.1037/a0021897](https://doi.org/10.1037/a0021897)] [Medline: [21219045](https://pubmed.ncbi.nlm.nih.gov/21219045/)]
46. Bolier L, Haverman M, Westerhof G, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. *BMC Public Health* 2013 Feb 08;13:119 [FREE Full text] [doi: [10.1186/1471-2458-13-119](https://doi.org/10.1186/1471-2458-13-119)] [Medline: [23390882](https://pubmed.ncbi.nlm.nih.gov/23390882/)]
47. Carr A, Finnegan L, Griffin E, Cotter P, Hyland A. A randomized controlled trial of the say yes to life (SYTL) positive psychology group psychotherapy program for depression: an interim report. *J Contemp Psychother* 2016 Sep 28;47(3):153-161 [FREE Full text] [doi: [10.1007/s10879-016-9343-6](https://doi.org/10.1007/s10879-016-9343-6)]
48. Hendriks T, Schotanus-Dijkstra M, Hassankhan A, de Jong J, Bohlmeijer E. The efficacy of multi-component positive psychology interventions: a systematic review and meta-analysis of randomized controlled trials. *J Happiness Stud* 2019 Feb 11;21(1):357-390 [FREE Full text] [doi: [10.1007/s10902-019-00082-1](https://doi.org/10.1007/s10902-019-00082-1)]
49. Chakhssi F, Kraiss J, Sommers-Spijkerman M, Bohlmeijer E. The effect of positive psychology interventions on well-being and distress in clinical samples with psychiatric or somatic disorders: a systematic review and meta-analysis. *BMC Psychiatry* 2018 Jun 27;18(1):211 [FREE Full text] [doi: [10.1186/s12888-018-1739-2](https://doi.org/10.1186/s12888-018-1739-2)] [Medline: [29945603](https://pubmed.ncbi.nlm.nih.gov/29945603/)]
50. Geerling B, Kraiss J, Kelders S, Stevens A, Kupka R, Bohlmeijer E. The effect of positive psychology interventions on well-being and psychopathology in patients with severe mental illness: a systematic review and meta-analysis. *J Positive Psychol* 2020 Jul 10;15(5):572-587 [FREE Full text] [doi: [10.1080/17439760.2020.1789695](https://doi.org/10.1080/17439760.2020.1789695)]
51. Celano C, Gomez-Bernal F, Mastromauro C, Beale E, DuBois C, Auerbach R, et al. A positive psychology intervention for patients with bipolar depression: a randomized pilot trial. *J Ment Health* 2020 Feb;29(1):60-68 [FREE Full text] [doi: [10.1080/09638237.2018.1521942](https://doi.org/10.1080/09638237.2018.1521942)] [Medline: [30366513](https://pubmed.ncbi.nlm.nih.gov/30366513/)]
52. Farquharson L, MacLeod AK. A brief goal-setting and planning intervention to improve well-being for people with psychiatric disorders. *Psychother Psychosom* 2014;83(2):122-124 [FREE Full text] [doi: [10.1159/000356332](https://doi.org/10.1159/000356332)] [Medline: [24457329](https://pubmed.ncbi.nlm.nih.gov/24457329/)]
53. Painter J, Mote J, Peckham A, Lee E, Campellone T, Pearlstein J, et al. A positive emotion regulation intervention for bipolar I disorder: treatment development and initial outcomes. *Gen Hosp Psychiatry* 2019;61:96-103 [FREE Full text] [doi: [10.1016/j.genhosppsy.2019.07.013](https://doi.org/10.1016/j.genhosppsy.2019.07.013)] [Medline: [31439286](https://pubmed.ncbi.nlm.nih.gov/31439286/)]
54. Kraiss J, Ten Klooster PM, Chrispijn M, Trompeter H, Stevens A, Neutel E, et al. B-positive: a randomized controlled trial of a multicomponent positive psychology intervention for euthymic patients with bipolar disorder - study protocol and intervention development. *BMC Psychiatry* 2018 Oct 17;18(1):335 [FREE Full text] [doi: [10.1186/s12888-018-1916-3](https://doi.org/10.1186/s12888-018-1916-3)] [Medline: [30333004](https://pubmed.ncbi.nlm.nih.gov/30333004/)]
55. Ten Klooster PM, Kupka RW, Chrispijn M, Stevens A. The effectiveness of a multicomponent positive psychology group intervention for euthymic patients with bipolar disorder: a pragmatic randomized controlled trial. *Data Archiving and Networked Services (DANS)*. 2021 Nov 26. URL: <https://www.narcis.nl/dataset/RecordID/oai%3Aeasy.dans.knaw.nl%3Aeasy-dataset%3A228487> [accessed 2022-09-05]
56. Bohlmeijer E, Westerhof G. A new model for sustainable mental health: integrating well-being into psychological treatment. In: *Making an Impact on Mental Health: The Applications of Psychological Research*. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2021.
57. Bohlmeijer E, Westerhof G. The model for sustainable mental health: future directions for integrating positive psychology into mental health care. *Front Psychol* 2021;12:747999 [FREE Full text] [doi: [10.3389/fpsyg.2021.747999](https://doi.org/10.3389/fpsyg.2021.747999)] [Medline: [34744925](https://pubmed.ncbi.nlm.nih.gov/34744925/)]
58. Bolier L, Abello K. Online positive psychological interventions: state of the art and future directions. In: *The Wiley Blackwell Handbook of Positive Psychological Interventions*. Hoboken, New Jersey, United States: Wiley; 2014.
59. Gander F, Proyer R, Ruch W. Positive psychology interventions addressing pleasure, engagement, meaning, positive relationships, and accomplishment increase well-being and ameliorate depressive symptoms: a randomized, placebo-controlled online study. *Front Psychol* 2016;7:686 [FREE Full text] [doi: [10.3389/fpsyg.2016.00686](https://doi.org/10.3389/fpsyg.2016.00686)] [Medline: [27242600](https://pubmed.ncbi.nlm.nih.gov/27242600/)]
60. Görges F, Oehler C, von Hirschhausen E, Hegerl U, Rummel-Kluge C. GET.HAPPY2 - user perspectives on an internet-based self-management positive psychology intervention among persons with and without depression: results from a retrospective survey. *J Clin Psychol* 2020 Jun;76(6):1030-1046 [FREE Full text] [doi: [10.1002/jclp.22886](https://doi.org/10.1002/jclp.22886)] [Medline: [31714609](https://pubmed.ncbi.nlm.nih.gov/31714609/)]
61. Wellenzohn S, Proyer R, Ruch W. Humor-based online positive psychology interventions: a randomized placebo-controlled long-term trial. *J Positive Psychol* 2016 Jan 28;11(6):584-594 [FREE Full text] [doi: [10.1080/17439760.2015.1137624](https://doi.org/10.1080/17439760.2015.1137624)]

62. Walsh S, Kaselionyte J, Taylor S, Priebe S. What might affect acceptability of online positive psychology interventions for depression: a qualitative study on patient expectations'. *BMC Psychiatry* 2018 Jul 27;18(1):240 [FREE Full text] [doi: [10.1186/s12888-018-1812-x](https://doi.org/10.1186/s12888-018-1812-x)] [Medline: [30053854](https://pubmed.ncbi.nlm.nih.gov/30053854/)]
63. Walsh S, Szymczynska P, Taylor S, Priebe S. The acceptability of an online intervention using positive psychology for depression: a qualitative study. *Internet Interv* 2018 Sep;13:60-66 [FREE Full text] [doi: [10.1016/j.invent.2018.07.003](https://doi.org/10.1016/j.invent.2018.07.003)] [Medline: [30206520](https://pubmed.ncbi.nlm.nih.gov/30206520/)]
64. Sergeant S, Mongrain M. Distressed users report a better response to online positive psychology interventions than nondistressed users. *Can Psychol / Psychologie canadienne* 2015 Aug;56(3):322-331 [FREE Full text] [doi: [10.1037/cap0000034](https://doi.org/10.1037/cap0000034)]
65. Lal S, Adair C. E-mental health: a rapid review of the literature. *Psychiatr Serv* 2014 Jan 01;65(1):24-32. [doi: [10.1176/appi.ps.201300009](https://doi.org/10.1176/appi.ps.201300009)] [Medline: [24081188](https://pubmed.ncbi.nlm.nih.gov/24081188/)]
66. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006 May 16;144(10):742-752 [FREE Full text] [doi: [10.7326/0003-4819-144-10-200605160-00125](https://doi.org/10.7326/0003-4819-144-10-200605160-00125)] [Medline: [16702590](https://pubmed.ncbi.nlm.nih.gov/16702590/)]
67. Beentjes TA, Goossens PJ, Vermeulen H, Teerenstra S, Nijhuis-van der Sanden MW, van Gaal BG. *BMC Health Serv Res* 2018 Dec 12;18(1):962 [FREE Full text] [doi: [10.1186/s12913-018-3767-5](https://doi.org/10.1186/s12913-018-3767-5)] [Medline: [30541536](https://pubmed.ncbi.nlm.nih.gov/30541536/)]
68. Safi S, Thiessen T, Schmailzl K. Acceptance and resistance of new digital technologies in medicine: qualitative study. *JMIR Res Protoc* 2018 Dec 04;7(12):e11072 [FREE Full text] [doi: [10.2196/11072](https://doi.org/10.2196/11072)] [Medline: [30514693](https://pubmed.ncbi.nlm.nih.gov/30514693/)]
69. Kujala S. User involvement: a review of the benefits and challenges. *Behav Inform Technol* 2003 Jan;22(1):1-16 [FREE Full text] [doi: [10.1080/01449290301782](https://doi.org/10.1080/01449290301782)]
70. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard H, Kelders S, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. *J Med Internet Res* 2011 Dec 05;13(4):e111 [FREE Full text] [doi: [10.2196/jmir.1672](https://doi.org/10.2196/jmir.1672)] [Medline: [22155738](https://pubmed.ncbi.nlm.nih.gov/22155738/)]
71. Kelders S. Design for engagement of online positive psychology interventions. In: *Positive Psychological Intervention Design and Protocols for Multi-Cultural Contexts*. Cham: Springer; 2019.
72. Shosha G. Employment of Colaizzi's strategy in descriptive phenomenology: a reflection of a researcher. *Eur Sci J* 2012;8(27) [FREE Full text] [doi: [10.19044/esj.2012.v8n27p%25p](https://doi.org/10.19044/esj.2012.v8n27p%25p)]
73. Denicoff K, Leverich G, Nolen W, Rush A, McElroy S, Keck P, et al. Validation of the prospective NIMH-Life-Chart Method (NIMH-LCM-p) for longitudinal assessment of bipolar illness. *Psychol Med* 2000 Nov;30(6):1391-1397 [FREE Full text] [doi: [10.1017/s0033291799002810](https://doi.org/10.1017/s0033291799002810)] [Medline: [11097079](https://pubmed.ncbi.nlm.nih.gov/11097079/)]
74. Bohlmeijer E, Hulsbergen M. *Using Positive Psychology Every Day Learning How to Flourish*. London: Routledge; 2018.
75. Iasiello M, van Agteren J, Keyes C, Cochrane E. Positive mental health as a predictor of recovery from mental illness. *J Affect Disord* 2019 May 15;251:227-230 [FREE Full text] [doi: [10.1016/j.jad.2019.03.065](https://doi.org/10.1016/j.jad.2019.03.065)] [Medline: [30927584](https://pubmed.ncbi.nlm.nih.gov/30927584/)]
76. Tse S, Murray G, Chung K, Davidson L, Ng K, Yu C. Exploring the recovery concept in bipolar disorder: a decision tree analysis of psychosocial correlates of recovery stages. *Bipolar Disord* 2014 Jun;16(4):366-377 [FREE Full text] [doi: [10.1111/bdi.12153](https://doi.org/10.1111/bdi.12153)] [Medline: [24261315](https://pubmed.ncbi.nlm.nih.gov/24261315/)]
77. Keyes C, Dhingra S, Simoes E. Change in level of positive mental health as a predictor of future risk of mental illness. *Am J Public Health* 2010 Dec;100(12):2366-2371 [FREE Full text] [doi: [10.2105/ajph.2010.192245](https://doi.org/10.2105/ajph.2010.192245)]
78. Gruber J, Culver JL, Johnson SL, Nam JY, Keller KL, Ketter TA. Do positive emotions predict symptomatic change in bipolar disorder? *Bipolar Disord* 2009 May;11(3):330-336 [FREE Full text] [doi: [10.1111/j.1399-5618.2009.00679.x](https://doi.org/10.1111/j.1399-5618.2009.00679.x)] [Medline: [19419390](https://pubmed.ncbi.nlm.nih.gov/19419390/)]
79. Harvey A, Talbot L, Gershon A. Sleep disturbance in bipolar disorder across the lifespan. *Clin Psychol (New York)* 2009 Jun;16(2):256-277 [FREE Full text] [doi: [10.1111/j.1468-2850.2009.01164.x](https://doi.org/10.1111/j.1468-2850.2009.01164.x)] [Medline: [22493520](https://pubmed.ncbi.nlm.nih.gov/22493520/)]
80. Proudfoot J, Whitton A, Parker G, Doran J, Manicavasagar V, Delmas K. Triggers of mania and depression in young adults with bipolar disorder. *J Affect Disord* 2012 Dec 20;143(1-3):196-202 [FREE Full text] [doi: [10.1016/j.jad.2012.05.052](https://doi.org/10.1016/j.jad.2012.05.052)] [Medline: [22884233](https://pubmed.ncbi.nlm.nih.gov/22884233/)]
81. Gilbert P, McEwan K, Matos M, Ravis A. Fears of compassion: development of three self-report measures. *Psychol Psychother* 2011 Sep;84(3):239-255 [FREE Full text] [doi: [10.1348/147608310X526511](https://doi.org/10.1348/147608310X526511)] [Medline: [22903867](https://pubmed.ncbi.nlm.nih.gov/22903867/)]
82. Rosenfarb I, Becker J, Khan A, Mintz J. Dependency and self-criticism in bipolar and unipolar depressed women. *Br J Clin Psychol* 1998 Nov;37(4):409-414 [FREE Full text] [doi: [10.1111/j.2044-8260.1998.tb01398.x](https://doi.org/10.1111/j.2044-8260.1998.tb01398.x)] [Medline: [9856294](https://pubmed.ncbi.nlm.nih.gov/9856294/)]
83. Nitzburg G, Russo M, Cuesta-Diaz A, Ospina L, Shanahan M, Perez-Rodriguez M, et al. Coping strategies and real-world functioning in bipolar disorder. *J Affect Disord* 2016 Jul 01;198:185-188 [FREE Full text] [doi: [10.1016/j.jad.2016.03.028](https://doi.org/10.1016/j.jad.2016.03.028)] [Medline: [27017375](https://pubmed.ncbi.nlm.nih.gov/27017375/)]
84. Geerling B, Kelders S, Kupka R, Stevens A, Bohlmeijer E. How to make online mood-monitoring in bipolar patients a success? A qualitative exploration of requirements. *Int J Bipolar Disord* 2021 Dec 01;9(1):39 [FREE Full text] [doi: [10.1186/s40345-021-00244-2](https://doi.org/10.1186/s40345-021-00244-2)] [Medline: [34851456](https://pubmed.ncbi.nlm.nih.gov/34851456/)]
85. Carr A, Cullen K, Keeney C, Canning C, Mooney O, Chinseallaigh E, et al. Effectiveness of positive psychology interventions: a systematic review and meta-analysis. *J Positive Psychol* 2020 Sep 10;16(6):749-769 [FREE Full text] [doi: [10.1080/17439760.2020.1818807](https://doi.org/10.1080/17439760.2020.1818807)]

86. Borghouts J, Eikeij E, Mark G, De Leon C, Schueller S, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. *J Med Internet Res* 2021 Mar 24;23(3):e24387 [[FREE Full text](#)] [doi: [10.2196/24387](https://doi.org/10.2196/24387)] [Medline: [33759801](https://pubmed.ncbi.nlm.nih.gov/33759801/)]
87. van Bruinessen IR, van Weel-Baumgarten EM, Snippe H, Gouw H, Zijlstra J, van Dulmen S. Active patient participation in the development of an online intervention. *JMIR Res Protoc* 2014 Nov 06;3(4):e59 [[FREE Full text](#)] [doi: [10.2196/resprot.3695](https://doi.org/10.2196/resprot.3695)] [Medline: [25379679](https://pubmed.ncbi.nlm.nih.gov/25379679/)]
88. Preece J, Sharp H, Rogers Y. *Interaction Design: Beyond Human-Computer Interaction*. Hoboken, New Jersey: Wiley; 2015.
89. Kelders S, Pots W, Oskam M, Bohlmeijer E, van Gemert-Pijnen JE. Development of a web-based intervention for the indicated prevention of depression. *BMC Med Inform Decis Mak* 2013 Feb 20;13:26 [[FREE Full text](#)] [doi: [10.1186/1472-6947-13-26](https://doi.org/10.1186/1472-6947-13-26)] [Medline: [23425322](https://pubmed.ncbi.nlm.nih.gov/23425322/)]
90. Litosseliti L. *Using Focus Groups in Research*. London, United Kingdom: Bloomsbury Publishing; 2003.

## Abbreviations

**BD:** bipolar disorder  
**CeHRes:** Center for eHealth and Disease Management  
**CHIME:** connectedness, hope, identity, meaning, and empowerment  
**FG:** focus group  
**FGD:** focus group discussion  
**LCM:** Life Chart Method  
**OPPI:** online positive psychology intervention  
**PP:** positive psychology  
**PPI:** positive psychology intervention  
**PPT:** paper prototyping  
**PT:** pilot test  
**RPT:** rapid prototyping  
**VS:** value specification

*Edited by A Mavragani; submitted 11.05.22; peer-reviewed by E Kawamoto, M Brown; comments to author 20.07.22; revised version received 04.08.22; accepted 09.08.22; published 19.09.22.*

*Please cite as:*

Geerling B, Kelders SM, Stevens AWMM, Kupka RW, Bohlmeijer ET  
*A Web-Based Positive Psychology App for Patients With Bipolar Disorder: Development Study*  
*JMIR Form Res* 2022;6(9):e39476  
URL: <https://formative.jmir.org/2022/9/e39476>  
doi: [10.2196/39476](https://doi.org/10.2196/39476)  
PMID: [35946327](https://pubmed.ncbi.nlm.nih.gov/35946327/)

©Bart Geerling, Saskia M Kelders, Anja W M M Stevens, Ralph W Kupka, Ernst T Bohlmeijer. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 19.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Resilience in Web-Based Mental Health Communities: Building a Resilience Dictionary With Semiautomatic Text Analysis

Yong-Bin Kang<sup>1</sup>, PhD; Anthony McCosker<sup>1,2</sup>, PhD; Peter Kamstra<sup>2</sup>, PhD; Jane Farmer<sup>2</sup>, PhD

<sup>1</sup>Australian Research Council (ARC) Centre of Excellence for Automated Decision-Making and Society (ADM+S), Swinburne University of Technology, Victoria, Australia

<sup>2</sup>Social Innovation Research Institute, Swinburne University of Technology, Victoria, Australia

**Corresponding Author:**

Yong-Bin Kang, PhD

Australian Research Council (ARC) Centre of Excellence for Automated Decision-Making and Society (ADM+S)

Swinburne University of Technology

John St

Hawthorn

Victoria, 3122

Australia

Phone: 61 3 9214 5904

Email: [ykang@swin.edu.au](mailto:ykang@swin.edu.au)

## Abstract

**Background:** Resilience is an accepted strengths-based concept that responds to change, adversity, and crises. This concept underpins both personal and community-based preventive approaches to mental health issues and shapes digital interventions. Online mental health peer-support forums have played a prominent role in enhancing resilience by providing accessible places for sharing lived experiences of mental issues and finding support. However, little research has been conducted on whether and how resilience is realized, hindering service providers' ability to optimize resilience outcomes.

**Objective:** This study aimed to create a resilience dictionary that reflects the characteristics and realization of resilience within online mental health peer-support forums. The findings can be used to guide further analysis and improve resilience outcomes in mental health forums through targeted moderation and management.

**Methods:** A semiautomatic approach to creating a resilience dictionary was proposed using topic modeling and qualitative content analysis. We present a systematic 4-phase analysis pipeline that preprocesses raw forum posts, discovers core themes, conceptualizes resilience indicators, and generates a resilience dictionary. Our approach was applied to a mental health forum run by SANE (Schizophrenia: A National Emergency) Australia, with 70,179 forum posts between 2018 and 2020 by 2357 users being analyzed.

**Results:** The resilience dictionary and taxonomy developed in this study, reveal how resilience indicators (ie, "social capital," "belonging," "learning," "adaptive capacity," and "self-efficacy") are characterized by themes commonly discussed in the forums; each theme's top 10 most relevant descriptive terms and their synonyms; and the relatedness of resilience, reflecting a taxonomy of indicators that are more comprehensive (or compound) and more likely to facilitate the realization of others. The study showed that the resilience indicators "learning," "belonging," and "social capital" were more commonly realized, and "belonging" and "learning" served as foundations for "social capital" and "adaptive capacity" across the 2-year study period.

**Conclusions:** This study presents a resilience dictionary that improves our understanding of how aspects of resilience are realized in web-based mental health forums. The dictionary provides novel guidance on how to improve training to support and enhance automated systems for moderating mental health forum discussions.

(*JMIR Form Res* 2022;6(9):e39013) doi:[10.2196/39013](https://doi.org/10.2196/39013)

**KEYWORDS**

resilience dictionary; mental health; peer-support forum; topic modeling; text analysis; content moderation

## Introduction

### Background

*Mental health* is fundamental to our individual and collective capabilities of thinking and interacting with each other and enjoying life [1]. Mental health challenges have been increasing worldwide, with consequent negative impacts on social and economic prosperity [2-4]. A 2017 survey [5] found that approximately 1 in 7 people (equivalent to 1 billion people) worldwide have experienced mental illness. In Australia, government statistical reports show that approximately 20% of people (4.8 million) experience mental health issues such as anxiety-related conditions and depression [6]. From 2020 to 2021 during the COVID-19 pandemic, 3.4 million Australians aged 16 to 85 years (17% of that age group) sought help from a mental health professional [7]. With its lockdowns and social distancing requirements, the COVID-19 pandemic is associated with what some have called an “unprecedented mental health crisis” [8], accelerating the need for community and digital interventions [9].

Reaching out to others for support is a vital part of enhancing resilience [10]. On this basis, it has been shown that *online peer-support mental health forums* (eg, those run in Australia by the charities Beyond Blue [11], SANE [Schizophrenia: A National Emergency] Australia [12], and ReachOut [13]) play increasingly important roles in establishing social connections, sharing knowledge and experiences, and providing emotional support among people with lived experiences of mental illness [4,14]. Such forums complement publicly funded and private health services by enabling firsthand access to people with shared experiences, advice, and guidance. Within forums, specially trained staff, volunteers, and untrained peers provide a supportive and safe space to talk and be heard [15,16]. Evidence shows that the benefits include the following [17]: (1) building safe and trusting relationships, (2) ensuring values of mutuality and reciprocity, (3) promoting validation and application of experiential knowledge, (4) enabling peers to exercise leadership through peer support, and (5) empowering peers to discover and make use of their own strengths. As complements to clinical and telehealth services, there remains a need to better understand the strengths of peer-support forums and optimize their management and moderation.

*Resilience* is an accepted strengths-based concept rather than a deficit or harms-based concept. Generating resilience is a regularly suggested approach to addressing mental health issues in community settings and falls within a preventive model of health care [8,18,19]. Resilience is understood variously across several disciplines as an individual and collective capability to adapt in the face of adversity, trauma, tragedy, threats, or even significant sources of stress [20]. In response to challenges, resilient individuals and communities can draw on psychosocial resources—personal and collective—to cope and adapt [4]. As a result, an individual’s resilience can positively affect their mental health. However, more research is needed to understand the characteristics of resilience and how it is realized through interactions within dedicated web-based mental health support settings, such as peer-support forums. This research contributes

a method, as well as findings, that can help forum service providers demonstrate their outcomes and tailor their moderation practices to optimize resilience-building interactions.

### Prior Work

Despite the potential utility of online mental health peer-support forums for building resilience, *whether and how resilience is realized* by participants has received little research attention. Where resilience or strengths-based resources are the focus of existing studies, approaches fall into two streams: (1) qualitative analysis using user surveys and (2) qualitative and quantitative analysis of web-based forum data using either manual or larger-scale artificial intelligence techniques. The first stream is centered on manually identifying aspects of resilience and their reciprocal associations with mental health from web-based user surveys during particular events (eg, the COVID-19 pandemic [9,19]) or from certain groups (eg, university students [18] or adolescents [21]). User survey data are less adept at showing how resilience is realized through forum interaction and discourse as it is cross-sectional at points in time and asks specific questions created by survey designers, which might not resonate with forum users’ perceptions and experiences.

The second research stream has evolved with recent advances in natural language processing (NLP) and machine learning. These methods have shown promise as a way of using forum content as data sets for exploring public health questions [22]. However, much of this work has been “risk” and “harms” focused rather than strengths focused. For example, a triaging system was used to assist web-based peer support by classifying forum messages into different risk levels based on how urgently moderators’ attention was needed [23]. A study [3] compared 2 web-based depression forums (Beyond Blue [11] and *r/depression* on Reddit [24]) using NLP techniques based on user sentiments and discussion topics. NLP techniques have also been used to identify which posts from web-based health forums (HealthBoards [25], Inspire [26], and HealthUnlocked [27]) are related to the COVID-19 pandemic based on conversations among mental health consumers. Other studies have applied NLP to detect sentiments and emotions (ie, fear, anger, sadness, and joy) [28-30] for depression diagnosis [31,32] and to understand grief processes [33] and stress [34]. As noted, these studies tend to target risks associated with mental ill health, leaving a gap in identifying strengths-based interactions in web-based data sets and how web-based services might help generate positive health outcomes.

### Aims of This Study

As mental health services seek to optimize and expand digital support interventions, new methods of analysis and monitoring of forum activities are needed. In this study, we harnessed NLP techniques to reveal indicators of resilience and how resilience is realized within online peer-support forums. This work provides new insights into how resilience is realized in forums by building a *resilience dictionary* that reveals the resilience themes discussed across 2 years of forum activity.

More specifically, the resilience dictionary we generated shows (1) the types of themes that characterize resilience and (2) the relationships between different resilience indicators with respect

to their realization. These insights reveal the broader (more abstract or common) and narrower (more specific) nature of specific resilience indicators. The resilience dictionary can be used by forum service providers and researchers to show that resilience is enabled by forums and how this occurs. It can help monitor changes in resilience realized by forum users over time and understand what prompts resilience changes, both within the forum and stimulated by external events. By establishing a method for quantifying resilience in unstructured text data, the resilience dictionary can help forum managers and creators to think more strategically about how to design the forums and maintain and moderate them more effectively. Revealing the characteristics of resilience and the extent to which it is realized over time can also help nonprofit organizations attract funding to sustain and further develop these forums.

To achieve our aims, we investigated the following research questions (RQs):

- RQ1: How can we characterize resilience indicators using topics (ie, major themes of posts) and their descriptive terms?
- RQ2: How are resilience indicators realized over time, and which are more dominant than the others?
- RQ3: How can we create a resilience dictionary that reflects the characteristics of each resilience indicator and the relatedness between the resilience indicators?

To address these RQs, we conducted a semiautomatic approach using methods that incorporate *topic modeling* (a type of statistical modeling in NLP) and exploit *human knowledge*.

To address RQ1, we explored optimal methods for discovering major thematic concepts, or *topics*, in forum data using topic modeling. Subsequently, using qualitative methods, we mapped these topics to the resilience indicators. The topics and their descriptive terms formed the controlled vocabulary of the resilience dictionary.

To investigate RQ2, we observed the *resilience prevalence* by analyzing the proportion of input forum data during the specified periods. We then analyzed the prevalence of each resilience indicator over time.

To explore RQ3, we created a resilience dictionary using the results of topics mapped to each resilience indicator. To identify the relatedness between resilience indicators in terms of their realization, we examined the co-occurrence patterns of resilience indicators from the input forums to automatically build a taxonomy of resilience indicators (ie, *resilience taxonomy*). This resilience taxonomy sheds light on whether resilience indicators are dependent on or independent of other resilience indicators.

For this work, we used data produced by SANE Australia's 2 forums. SANE forums represent one of the largest web-based mental health peer-support communities in Australia, where people aged  $\geq 18$  years with mental health issues can register and engage in support, training, and education. Users can interact with other people (ie, peers) experiencing similar mental health challenges through the forums.

To our knowledge, this study is the first to explore the construction of a resilience dictionary by examining relationships between topics and resilience indicators and analyzing the relatedness between resilience indicators from web-based mental health forums using NLP and human knowledge. The significance of this study lies in the following aspects. First, we present how strengths-based resilience develops in web-based mental health forums based on a wider spectrum of web-based mental health discussions. Prior studies have attempted to accomplish this; however, they were small in scale and limited by specific events (eg, the COVID-19 pandemic) [8,9,14] or by certain age groups [2,18,21]. Second, the resilience dictionary can provide evidence-based information that can help design and maintain better mental health support services, as well as improve responses to people seeking mental health support. In particular, this benefit can be valuable for nonprofit mental health organizations that often have difficulty securing and allocating resources to improve their services. Finally, this work shows how mental health services can be made more effective by using support forum data, the value of which is becoming increasingly recognized [22,30].

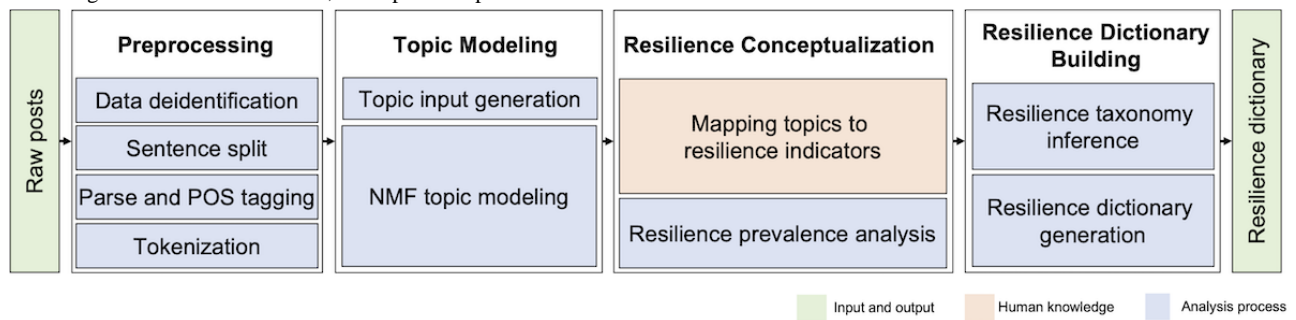
## Methods

### Overall Design

#### Overview

The 4 phases of our semiautomatic approach to creating a resilience dictionary are shown in [Figure 1](#). Forum posts were the *first* preprocess to ensure deidentification and perform tokenization to generate meaningful terms. *Second*, topic modeling was used to discover the key topics (ie, themes) discussed in the posts. *Third*, the topics were mapped to a framework of resilience indicators using thematic coding processes. *Fourth*, a resilience taxonomy was established based on the co-occurrence analysis of resilience indicators from the posts. A resilience dictionary was then generated by integrating the resilience taxonomy with all the outcomes of the previous 4 phases. In the following sections, we present a detailed description of the phases after introducing the data set and the nature of the resilience indicators used in the study.

**Figure 1.** The 4-phase analysis pipeline that preprocesses forum posts, finds topics, conceptualizes resilience indicators, and builds a resilience dictionary. NMF: nonnegative matrix factorization; POS: parts of speech.



### ***SANE Forum Data***

SANE Australia is a national mental health charity that supports Australians affected by mental illness by providing peer-support services, training, and counseling. SANE data were collected from two web-based forums: (1) the “Lived Experience Forum” and (2) the “Friend, Family and Carers Forum.” The former is for people who live with complex mental health issues, whereas the latter is for their supporters and carers. SANE forums are moderated by mental health practitioners and aim to offer a safe and anonymous space on the web for discussing mental health and related issues [35]. Forums provide participants with social connections, allow the sharing of feelings and emotions, and provide a source of information generated through conversations. Forum discussions are intended to help consumers affected by mental illness explore positive pathways to address issues affecting their lives. According to the SANE annual report for 2020, there were 21,041 registered forum members, and 151,137 individual Australians used the forums in 2020 [35].

We obtained a national sample of posts from the SANE forums from 2018 Q3 (starting July 2018) to 2020 Q4 (ending December 2020). In total, 70,179 posts by 2357 users were obtained and used in this study in collaboration with the SANE forum managers.

### ***Resilience Indicators***

Drawing on the resilience framework proposed by Berkes and Ross [36], which combines psychological and community development approaches, we identified common components of personal and community resilience. Berkes and Ross [36] devised a conceptual framework of resilience pertaining to people’s lives in communities, allowing the concept to account

for the wider ecosystem of factors affecting an individual’s mental health and well-being. Although there is a range of approaches to resilience, this framework is useful for understanding how resilience is realized by members of a forum, given their location in both physical on-the-ground communities and the web-based peer-based community of the forums.

The adopted resilience approach is premised on the understanding that people and communities that are resilient require access to a set of personal, interpersonal, or community resources. On the basis of research evidence, these involve *learning* or access to new knowledge, information, and skills, leading to an increased capability of dealing with change [37]; *social capital* or access to networks of people and the support, trust, and integration they can foster; a sense of *belonging*, including belonging to groups and places, acceptance as part of a group, and the processes that lead to identity formation [38]; *adaptive capacity* or using resources to enable adaptation and positive behavior change [39]; and *self-efficacy* or being able to self-organize or work toward feeling in a controlled state [36]. The definitions of these indicators and how they are conceptually realized in the context of mental health are shown in Table 1. In this study, the generation of the resilience dictionary was focused on and built from these 5 indicators.

Theoretically, we expect to see some variation in the way each indicator of resilience manifests in the web-based discussion context (in comparison with other contexts) and differences in their prevalence [36]. This means that some indicators should act as generators or resources, and others act as attributes or characteristics of resilience capacity, such as adaptive capacity or instances of self-efficacy. These dependencies were tested through RQ3, with the application of the taxonomic analysis described in phase 4.

**Table 1.** Resilience indicators, their definitions, and their conceptual realization in web-based forums.

Resilience indicators	Definition	Realization in community forums
Social capital [38]	Access to social networks and the support, trust, and social integration they can foster	<ul style="list-style-type: none"> <li>• Formation of social connections that individuals could draw on for confidence and support</li> <li>• Expressing a willingness to share stories of similar experiences with mental health issues</li> </ul>
Belonging [38]	Belonging to people and places, acceptance as part of a group, and the processes that lead to identifying formation	<ul style="list-style-type: none"> <li>• Introducing themselves to others</li> <li>• Sharing stories and lived experiences with those experiencing similar challenges and participating at their own pace</li> </ul>
Learning [37]	Access to knowledge, information, and skill development	<ul style="list-style-type: none"> <li>• Sharing applied and experiential knowledge with others</li> <li>• Offering advice on how to practice good mental health and strategies on how to navigate mental health services</li> </ul>
Adaptive capacity [39]	Use of resources that enable adaptation and positive behavior change	<ul style="list-style-type: none"> <li>• Self-reporting their own activity or behavior changes when responding to advice or information on the forum</li> </ul>
Self-efficacy [36]	Ability to self-organize or work toward feeling in a controlled state	<ul style="list-style-type: none"> <li>• Communicating feelings of control in the aspects of their life or condition</li> <li>• Supporting others by keeping them accountable for their changed behaviors or positive lifestyle choices that made them feel in control</li> </ul>

### Ethics Approval

The university ethics committee approved this study (R/2019/033). In addition, we adhered to SANE's ethics and data governance policies throughout, which included establishing a data-sharing agreement, anonymizing posts to protect forum users' identities, and applying a data security protocol.

### Phase 1: Preprocessing of Forum Posts

In the preprocessing phase, the goal was to break up each post into smaller basic units of meaning. User identifiers (IDs or names) were removed to deidentify the posts. Posts were then split into sentences to perform part-of-speech tagging and remove stop words (words such as *is*, *are*, and *the*, which do not carry useful information). The remaining terms were converted into their lemmatized forms to group the inflected forms of words together. Posts that were too short to add value to topic modeling (<5 terms) were removed. Finally, the remaining terms were used as the input to the next "topic modelling" phase.

### Phase 2: Topic Modeling

#### Overview

Topic modeling has gained attention because of its capacity to discover latent thematic topics in a large corpus of text documents [40,41]. A topic is a specific, recognizable theme defined by a cohesive set of terms representing the characteristics of that theme. Nonnegative matrix factorization (NMF) [42] was chosen as a method because of its effectiveness in discovering topics in short text documents and its ability to discover both broad and specific topics. As a method of analyzing forum posts, NMF can help identify both broad themes discussed often over time and more specific themes that might relate to periodic events [43].

A key idea of NMF is to decompose a term-document matrix  $n$  by  $m$   $A$  ( $n$  is the number of forum posts, and  $m$  is the number of terms) into 2 nonnegative submatrices,  $n$  by  $k$   $W$  ( $k$  is the number of topics) and  $k$  by  $m$   $H$ , such that  $A$  is approximated by the multiplication of  $W$  and  $H$ , denoted as  $A \approx W \times H$ . In  $A$ , the weight of a term in a document is measured by a well-defined method, the term frequency-inverse document frequency weighting scheme [44], via term frequency and rarity. The matrix  $W$  represents the document membership weights over the topics. Each row denotes a document, and each column corresponds to a topic. Sorting the values of a column (topic) provides the ranking of the most relevant documents for the column. The matrix  $H$  contains the term weights relative to each of the topics. A row defines a topic. Sorting the values on each row provides the ranking of the most relevant terms (descriptors) of each topic.

#### Topic Input Generation

After preprocessing, each post was represented by a list of terms. Before applying topic modeling, we performed 2 steps to generate topic inputs. In the first step, the posts were split into disjoint collections based on the time duration of a "quarter" of a year. This collection period was the best fit for identifying granular topics (ie, neither too broad nor too specific). This generated 10 post collections from 2018 Q3 to 2020 Q4. The second step involved building a term-document matrix  $A$  using the term frequency-inverse document frequency weighting scheme from each collection.

#### Two-Layered NMF Topic Modeling

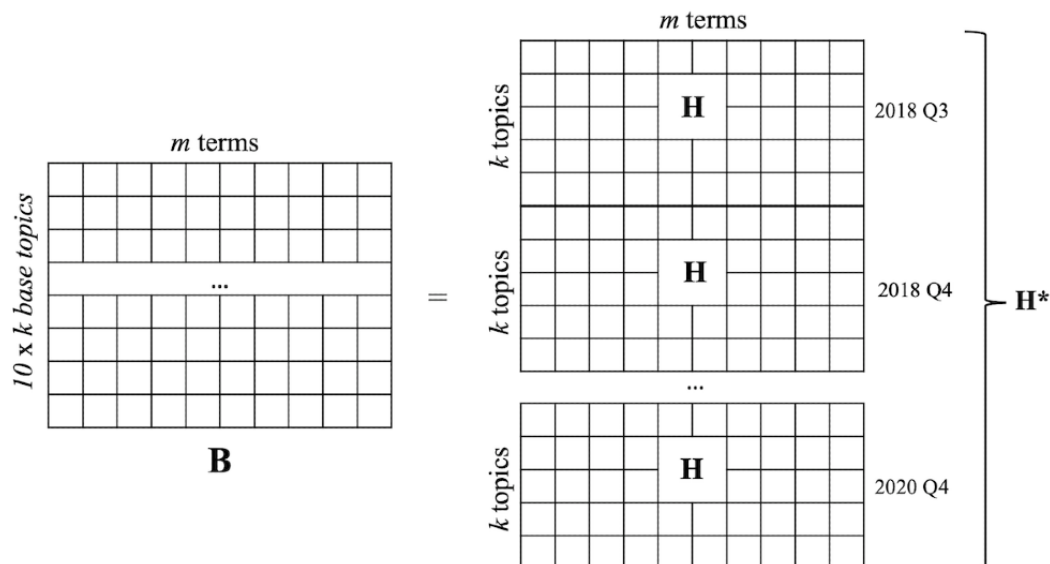
The 2-layered NMF approach [45] was applied to term-document matrices (denoted by  $A^*$ ) to discover both broader (longer-lived) and more specific (shorter-lived) topics. This process extracts broader topics observed across the entire time frame (2018 Q3 to 2020 Q4), as well as specific topics in each quarterly period, generating 2 topic layers. In the first layer, we find  $k$  topics, called *base topics*, from each term-document



matrix  $A$  of  $A^*$  using NMF, where  $k$  is a user-specified parameter. The output of this layer is a set of 2 matrices,  $W$  and  $H$ , from all disjoint post collections, denoted by  $W^*$  and  $H^*$ , respectively. In the second layer, we identified another type of topic, called *ensemble topics*, by analyzing similarities and variations among all the base topics generated from 2018 Q3 to 2020 Q4. A base topic-term matrix  $B$  was created by stacking each matrix  $H$  from  $H^*$  (Figure 2). Each  $H$  comprises  $k$  topics

with their constituent terms. In  $B$ , each row corresponds to a base topic, and each column is a term from the original posts. Each entry in  $B$  shows the weight (or importance) of the association of a term with a base topic, where the weight was inferred by the first-layer NMF. As shown in Figure 2, the dimension of  $B$  is  $10 \times k$  by  $m$ , as  $H^*$ , which comprises 10 different matrices of  $H$  (from 2018 Q3 to 2020 Q4) and  $m$  terms.

**Figure 2.** Conceptual illustration of constructing the matrix  $B$  by stacking up the matrices of  $H^*$  generated by the first-layer nonnegative matrix factorization.



From  $B$ , we applied NMF to produce  $k'$  ensemble topics, where  $k'$  is a user-specified parameter, thus producing  $B \approx W' \times H'$ , where (1)  $W'$  denotes the base topic membership weights over the  $k'$  ensemble topics, and (2) the  $H'$  matrix shows the weights of terms for each ensemble topic. To produce the membership weights of the original posts over an ensemble topic, we calculated the following expression:

$$D \approx C \times W'^T \text{ (equation 1)}$$

Here,  $C$  is the document-base topic matrix made by stacking each matrix  $W$  of  $W^*$  in the same way as  $B$  was created from the matrices of  $H^*$ .  $D$  is the document-ensemble topic matrix and  $W'^T$  is the transpose of  $W'$ . Each row in  $D$  is an original post, and each column represents an ensemble topic.

### Phase 3: Resilience Conceptualization

#### Overview

In this section, we present how to conceptualize each resilience indicator using its relevant topics. In this context, conceptualization refers to the process of specifying the characteristics of resilience indicators based on their relevant discussion themes (ie, topics) from qualitative forum posts. First, we present how to characterize each indicator using the discovered topics. Second, we present a method to determine whether resilience indicators are time varying. This can provide additional insights into their dynamic characteristics over time.

#### Mapping Topics to Resilience Indicators

Here, our goal was to connect topics with relevant resilience indicators (ie, in relation to our predetermined resilience

framework discussed previously). For this, we conducted qualitative content analysis, where knowledge of what happens on the forum and its aims is used to assess the meanings of topics and link each topic to relevant indicators. As qualitative data, we analyzed the meaning of the top  $N$  ( $N$  is a user-specified parameter) descriptive terms for each topic. As complementary data, we also examined the posts most relevant to each topic to analyze the meaning of the topic more rigorously in relation to the indicators of resilience. This content analysis was deductively performed using the framework of predefined resilience indicators. Note that a topic can be mapped to  $>1$  resilience indicator depending on the thematic analysis. We followed a recent work showing the benefits of combining topic modeling with qualitative methods, particularly in the interpretation and contextualization phases of the analysis [46]. Here, our focus was on gaining an understanding of the nature of topics and their associations with resilience indicators based on domain knowledge of the context and posts that contain them most frequently. Topics cannot be meaningfully interpreted based only on the top  $N$  words as topics are not independent of the posts in which they appear.

The two types of qualitative data generated and analyzed were as follows.

First, the top  $N$  terms that best describe each ensemble topic (henceforth simply, “topic”) were identified from the topic-term matrix  $H'$  generated from the matrix  $B$  (recall  $B \approx W' \times H'$  presented in phase 2; Figure 2). Note that each row in  $H'$  indicates the weights of all terms with respect to a topic. We generated these top  $N$  terms by ranking the weights. Examining

the collective meaning of these terms informed the mapping process of each topic to the relevant resilience indicators. For the value of  $N$ , we used 15.

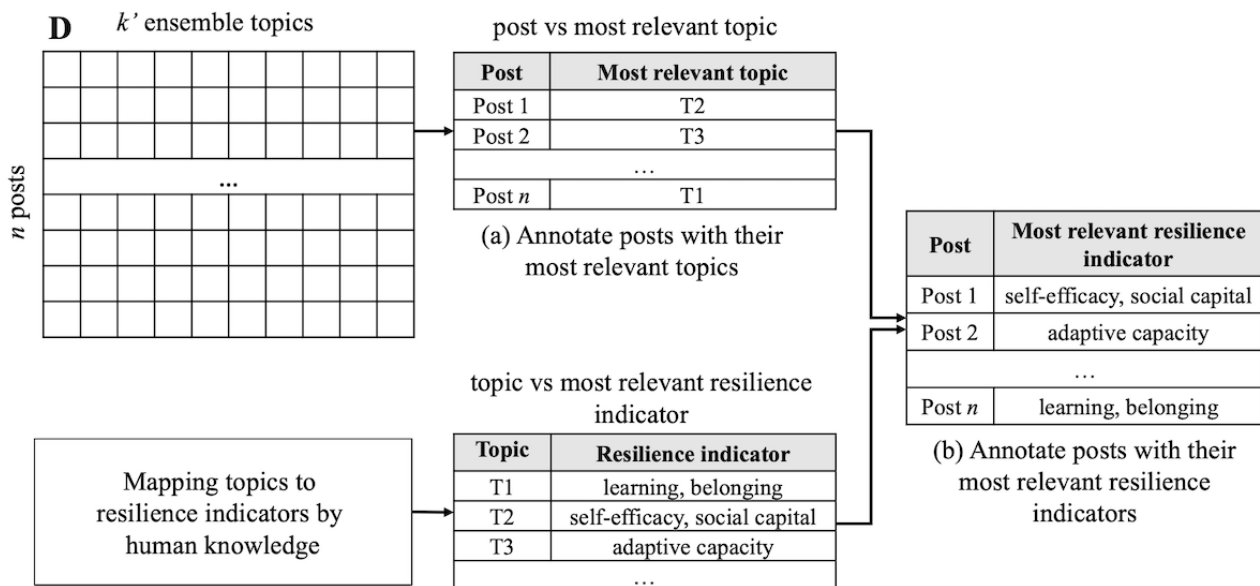
Second, we associated each topic with its most relevant original topic. This association was identified from matrix  $D$  (equation 1), which signifies the membership weights of the original posts for the discovered topics. By ranking these weights, we identified the top  $M$  most relevant posts for each topic. In our study, we chose to use 20 as the value of  $M$ . By conducting a content analysis of these posts, we gained insights into the range of words and ideas surrounding each topic.

**Resilience Prevalence Analysis**

Resilience prevalence analysis can identify the dynamics of resilience indicators in terms of their realization over time. We estimated resilience prevalence as the proportion of the original

posts associated with their most relevant resilience indicators. Using this analysis, we determined which resilience indicators were significantly or vaguely realized at a particular period. Thus, a resilience realization pattern over time was estimated using resilience prevalence analysis. To analyze resilience prevalence, it is essential to *annotate* each post with its most relevant resilience indicator(s). This was achieved by the following 2 steps. First, recall that from the document-ensemble topic matrix  $D$  (equation 1), we identified the membership weight of each post over each topic. On the basis of the weights of each post over all topics, we annotated each post with the topic that had the highest weight; thus, this topic was seen as the most significant topic for the post. Second, we note that each topic is mapped to its relevant resilience indicators in the previous step (see the *Mapping Topics to Resilience Indicators* section). Figure 3 illustrates these 2 steps using artificial examples.

**Figure 3.** Illustration of the 2 steps used to annotate posts with their most relevant resilience indicators. The first step shows how to annotate each post with its most relevant topic from the document-ensemble topic matrix  $D$  in equation 1. The second step annotates each post with its most resilience indicators from the outcome of the first step and the mapped outcomes from topics to their most relevant resilience.



**Phase 4: Resilience Dictionary Building**

In this section, we present our motivation and approach for building a resilience taxonomy that is part of our resilience dictionary. We then elaborate on the definition and creation of the resilience dictionary.

**Resilience Taxonomy Inference**

Our objective in creating a resilience taxonomy was to improve our understanding of (1) the relatedness between resilience indicators in terms of their realization (ie, whether the indicators are independently realized or corealized) and (2) which indicators are likely to facilitate the realization of some other indicators. To achieve this objective, our approach exploits the realization co-occurrences of resilience indicators. These co-occurrences are observed through annotated posts (see the method in phase 3). The taxonomy can conceptualize the relationships of indicators in an automated way that informs the understanding of resilience realization relationships. In

general, the key relationship in a taxonomy is the “is-a” relationship [47]. The “is-a” relationship has a hierarchical structure, and thus, its transitivity is logically inferred by navigating the hierarchical relationships between entities in a taxonomy. Intuitively, within our context, the higher a resilience indicator is positioned in the resilience taxonomy, the broader or more abstract it is. A lower position indicates greater specificity. Given a hierarchical path in the taxonomy between 2 indicators, we can derive the lower-ranked indicators that most likely influence the realization of higher indicator(s). As a result, semantic knowledge derived from the taxonomy can facilitate the corealization analysis of resilience indicators.

To create a resilience taxonomy, our fundamental goal was to examine the co-occurrences of resilience indicators annotated in the original posts. In particular, we used the popular subsumption method [48,49] that has proven to be effective for taxonomy learning in NLP; that is, from such co-occurrence knowledge, we build that a resilience indicator  $x$  subsumes

another resilience indicator  $y$  if the posts annotated with  $y$  are a subset of the posts annotated with  $x$ . On the basis of this scheme, we can find taxonomic relations between resilience indicators.

### Resilience Dictionary Generation

By aggregating the findings to this point, we generated a resilience dictionary that represents (1) how resilience indicators

are characterized by particular topics, (2) how each topic is represented by specific descriptive terms and what terms are similar, and (3) what semantic relationships exist between resilience indicators. The resilience dictionary can improve the understanding of the semantic coverage of the meaning and realization of each resilience indicator. The dictionary comprises the features described in [Textbox 1](#).

**Textbox 1.** Features of the resilience dictionary.

#### Resilience dictionary features

- *Resilience indicator* indicates a resilience indicator.
- *Parent* indicates the direct parent resilience indicator for a given indicator identified in the resilience taxonomy.
- *Topic* provides a list of topics relevant to each resilience indicator determined using human knowledge.
- *Topic weight* indicates the average weight of each topic over time. To calculate this weight, we first found the topic most relevant to each post by identifying the highest membership weight of the post over topics from the document-ensemble topic matrix  $D$ . In a sense, this weight is also seen as the weight of the topic (the most significant topic) in the post. Second, we calculated the average weight of the most significant topics across all the posts. Thus, this feature represents the relative importance of each topic over all the topics.
- *Topic word* denotes the top 10 most descriptive terms of each topic.
- *Similar terms* characterize the synonyms of a topic term, aiming to increase the semantics of a topic term. As there is no existing synonym dictionary for web-based mental health communities, our approach exploits a machine learning technique called *word embedding*. Word embedding is used to identify synonyms of each topic term based on their co-occurrence with other terms in input forum posts. Note that word embedding tends to indicate similar words by identifying the nearest words that appear together in similar contexts. Thus, synonyms captured by word embedding may not be “pure” synonyms depending on the input context (eg, if 2 terms “good” and “bad” frequently co-occur together in similar contexts, these can be identified as synonyms). We used a model named word2vec [50] as a word-embedding model, which has demonstrated many advantages for analyzing the semantic analysis of words.

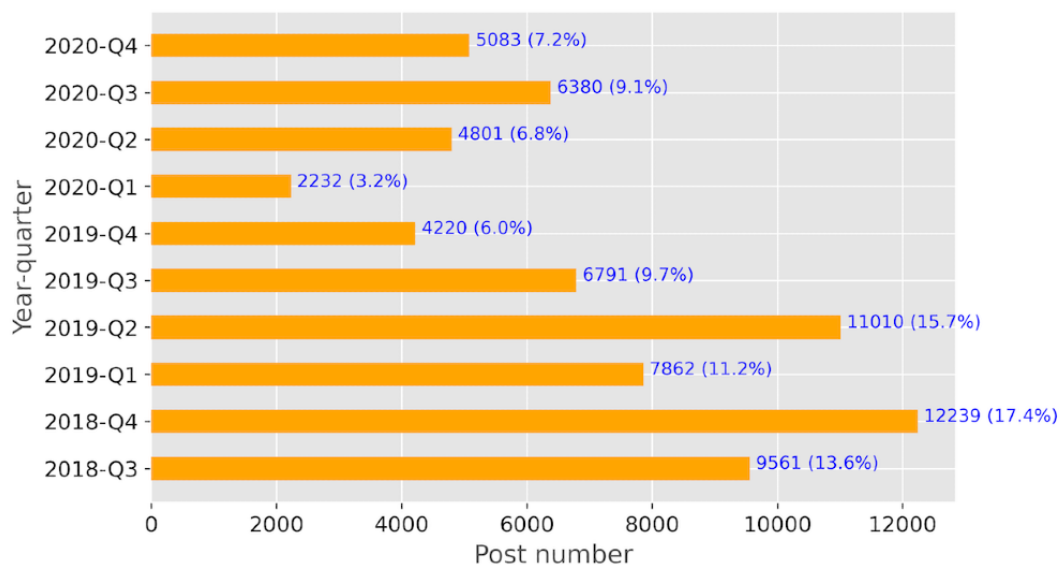
## Results

### Phase 1: Preprocessing of Mental Health Forum Posts

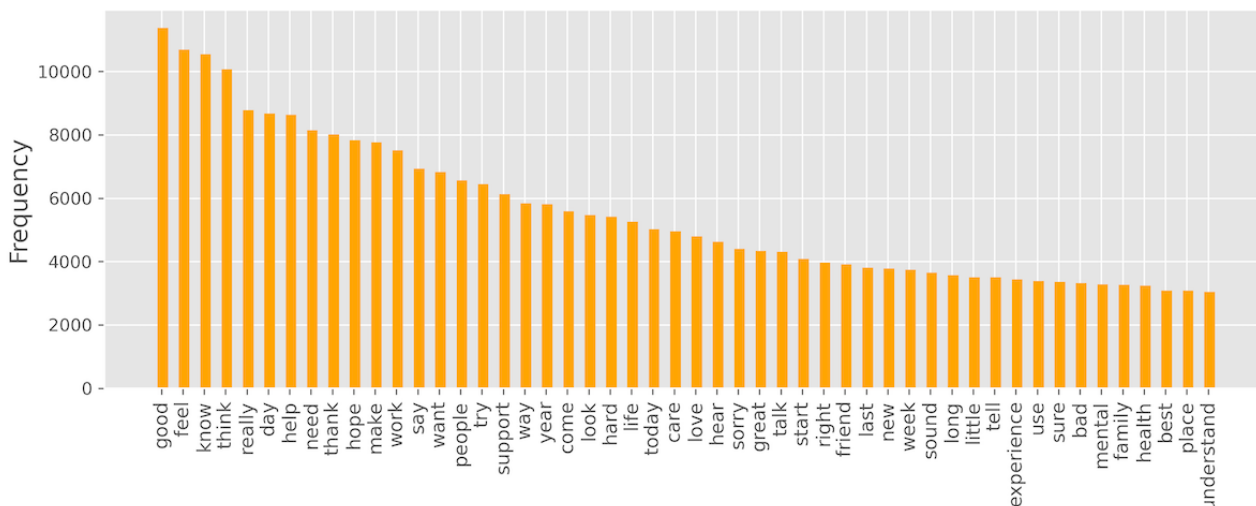
From the original 70,179 posts posted by 2357 users, [Figure 4](#) shows the number of posts by time quarter. As the input for topic modeling, we used 69.56% (48,819/70,179) of posts. Note that we removed 30.44% (21,360/70,179) of posts with a length of <5 words (ie, too short).

[Figure 5](#) shows the top 50 terms by word frequency in the post data. Most comprise a mixture of nouns (eg, *day*, *help*, *need*, *way*, and *year*), verbs (eg, *help*, *need*, *feel*, *know*, *think*, *want*, and *sorry*), and adjectives (eg, *good*, *hard*, *great*, and *little*), whereas adverbs were relatively less used (eg, *really*). In total, from the 48,819 posts, we extracted 14,938 terms after the preprocessing phase. The minimum, average, and maximum term counts of the posts were 6, 30, and 1182, respectively.

**Figure 4.** Summary of Schizophrenia: A National Emergency (SANE) forum posts in the sample 2018-Q3 to 2020-Q4.



**Figure 5.** The top 50 terms by word frequency observed from the 48,819 posts.

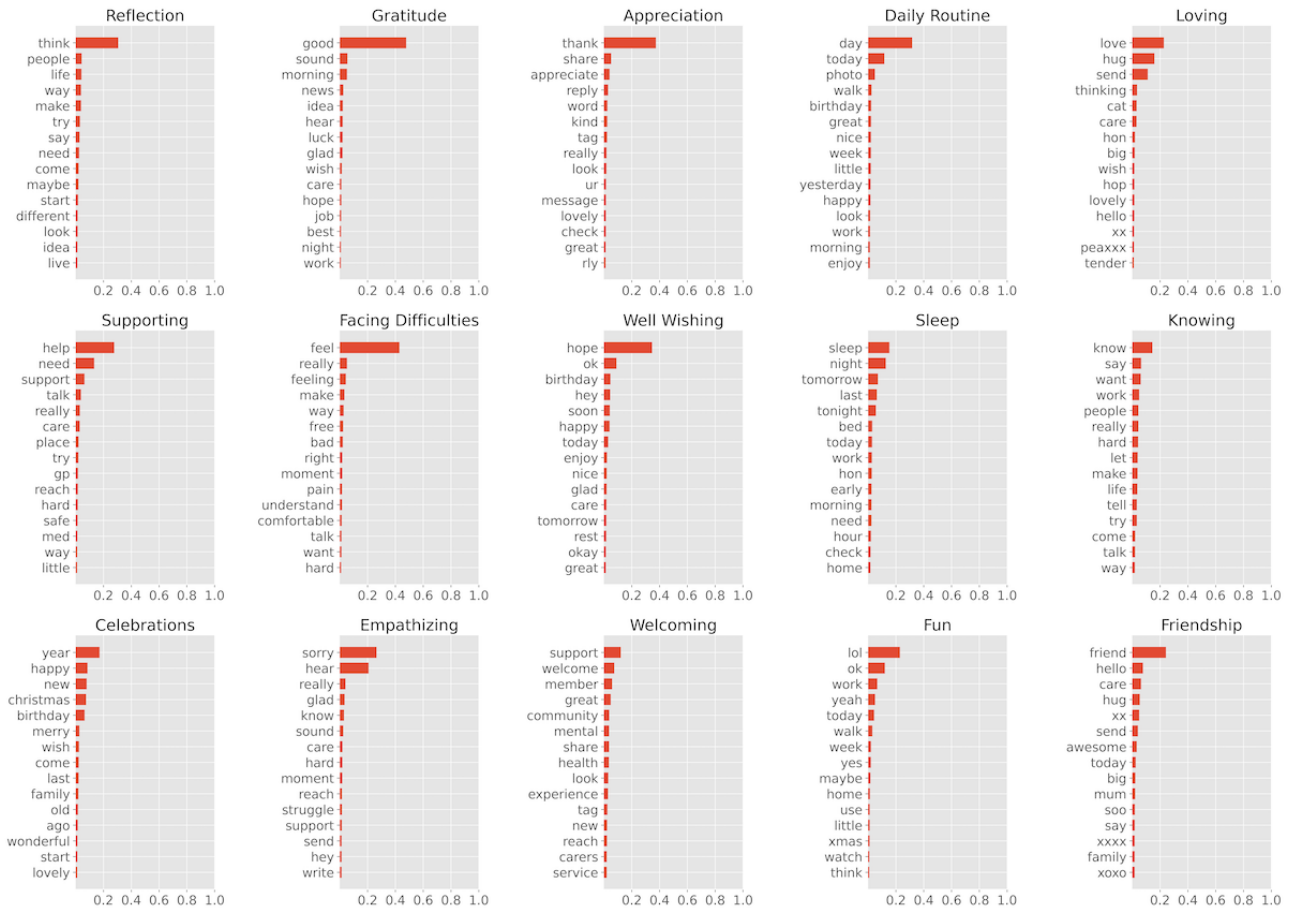


**Phase 2: Topic Modeling**

The 48,819 posts were split into 10 disjoint post collections based on quarters 2018 Q3 to 2020 Q4. From each collection, we discovered topics using NMF. To generate topics,  $k$  and  $k'$  must be provided, where  $k$  and  $k'$  determine the number of base and ensemble topics to be generated, respectively. There is no universal method of determining such numbers [43]. In our study, we used a fixed value of 10 as  $k$ , assuming that we equally discovered the 10 main topics being discussed from each collection. Initial experiments found that  $k=10$  (range was  $k=5-15$ ) produced an informative topic set that was neither too general nor too specific. To discover the  $k'$  ensemble topics, a more complex method was applied by varying  $k'$  from 10 to 20 in increments of 1 to choose an optimal number. As we did not know how many base topics were similar or different, we measured similarities and variances of the base topics over the entire timeline using a widely used metric, *topic coherence* [51].

This method measures the degree of similarity between the top  $N$  terms for each topic. From the perspective of topic coherence, a better-interpretable topic yields a higher average score for the topic coherence scores of all generated topics. To measure topic coherence, we used word embedding [43] to measure the similarity between the top  $N$  terms in a topic. This similarity was estimated using the cosine similarity between word-embedded vectors. For the value of  $N$ , we used 10, following the suggestion in the study by O’Callaghan et al [43]. We generated a word-embedding model from the entire post data using word2vec [50]. Finally, using topic coherence, we selected 15 as the optimal value for  $k'$ . Figure 6 shows the 15 ensemble topics generated. For each topic, the top 15 descriptive terms were ranked vertically (the first row is the most descriptive term) based on their weights (importance). The weight of each term in a topic is visualized by a “red bar,” where the sum of the weights of all terms in a topic is normalized to 1.

**Figure 6.** The 15 generated ensemble topics, where each topic is represented by its top 15 descriptive terms.



### Phase 3: Resilience Conceptualization

#### Results of Mapping Topics to Resilience Indicators

A qualitative thematic analysis following a previous work [52] was conducted by 4 researchers with expertise in qualitative mental health research (second, third, and fourth authors) to interpret the meaning of the topics and assign descriptive topic labels to them. These labels, as shown in Figure 6, describe each topic by finding the most suitable meaning conveyed by the top 15 words in the identified topics and the top 20 posts associated with that topic. First, the researchers familiarized themselves with the topic modeling results, independently noting the initial ideas that were used to provide a semantically representative label of topic words and posts. Subsequently, topic labels were compared and refined to achieve satisfactory interresearcher agreement. The final agreed topic labels were then attributed to  $\geq 1$  of the 5 resilience indicators by each researcher through a deductive process based on the established resilience theory. Finally, these groupings were reviewed by the team and refined for consistency (Table 2).

The identified topics reflect the forum’s aim to provide ongoing, accessible, peer-to-peer support for people experiencing mental health conditions. Although specific instances of mental ill health, such as depression and anxiety, were often discussed, they were not the focus of the identified topics; that is, although

some topics expressed mental health problems directly, such as “sleep” and “facing difficulties,” daily challenges and indicators of social connection and support were more commonly identified topics. Topics also have an important temporal dimension, which is an advantage of our methodology, which considers topics over time. The 2-year sample of posts and topic modeling based on quarters meant that annual events or celebrations such as Christmas, New Year’s Eve, and birthdays were commonly referenced, showing the importance of social connection over time as a feature of the forum. Most topics expressed the establishment and maintenance of interpersonal social connection: “welcoming,” “empathising,” “well wishing,” “supporting,” “loving,” “friendship,” and “appreciation.” A few topics were more self-directed: “gratitude,” “reflection,” and “knowing.” The latter is important but rarer indicators of adaptation and a sense of self that is necessary for resilience.

The researchers attributed topics to resilience indicators deductively, considering each topic’s top 15 words and associated top 20 posts. No assumptions were made regarding whether topics fit all resilience indicators and whether topics could be attributed to 1,  $>1$ , or no indicator. Strong agreement was achieved among the researchers, with some minor differences noted and discussed before an agreement was reached. The results of mapping topics to resilience indicators are shown in Table 2, with example posts.

**Table 2.** Result of mapping topics to resilience indicators.

Resilience indicator	Mapped topic	Example posts
Social capital	Gratitude, appreciation, daily routine, loving, supporting, well wishing, empathizing, fun, and friendship	<ul style="list-style-type: none"> <li>• “[...] you did good???? so you did do the dishes eventually. Just organised my bedside table and tidied backyard (courtyard), will get back to it in the morn, ta” [Gratitude]</li> <li>• “Wooooohhhooo thank you incredible human beings! Echoing x1000 [x’s] sentiment here” [Appreciation]</li> <li>• “The way to get help out of those cycles is to talk about them. Would it help to write them in a piece of paper and hand that over to the pdoc?” [Supporting]</li> </ul>
Belonging	Gratitude, appreciation, daily routine, loving, well wishing, celebrations, empathizing, welcoming, fun, and friendship	<ul style="list-style-type: none"> <li>• “Another welcome to the forum [...], it sounds like you’ve been through a lot so it’s understandable that you’re under so much stress.” [Welcoming]</li> <li>• “Happy New Year everyone in NSW, Vic, ACT and Tassie” [Celebrations]</li> <li>• “I get tired of my physical problems too but one step at a time my awesome friend, walking with you I am here for you all the way and so is [...]” [Friendship]</li> </ul>
Learning	Reflection, supporting, and knowing	<ul style="list-style-type: none"> <li>• “Listening to music does help. Need to learn to dismiss those unhelpful negative voices.” [Supporting]</li> <li>• “I’m trying so hard to see through my tears. it’s hard but I am determined to get there. Just knowing that you are here gives me strength x????” [Knowing]</li> </ul>
Adaptive capacity	Reflection	<ul style="list-style-type: none"> <li>• “Will be thinking of you through this process [...] and pray that the transition is as smooth as possible.” [Reflection]</li> </ul>
Self-efficacy	Facing difficulties and sleep	<ul style="list-style-type: none"> <li>• “Hi, I’m feeling very overwhelmed with anxiety atm. I’m trying to control it, but it feels like I’m a big shaking mess inside. I feel short of breath and my heart is racing” [Facing difficulties]</li> <li>• “Anxiety is off the charts. I’m sleep deprived as I got zero sleep last night and my mood is extremely low.” [Sleep]</li> </ul>

Most topics were attributed to the indicators “Social Capital” and “Belonging.” This aligns with the well-understood attributes of social media, which encourage and function through phatic communication, more so than informational or dialogic intents [53]. Within this broad sociopragmatic function of posts, we could see a range of topics that distinguished different methods of establishing and maintaining social connections. Topics associated with “Learning,” “Adaptive Capacity,” and “Self-efficacy” were rare, and these can be considered more explicit expressions of exhibited resilience, where change, adaptation, and coping become possible.

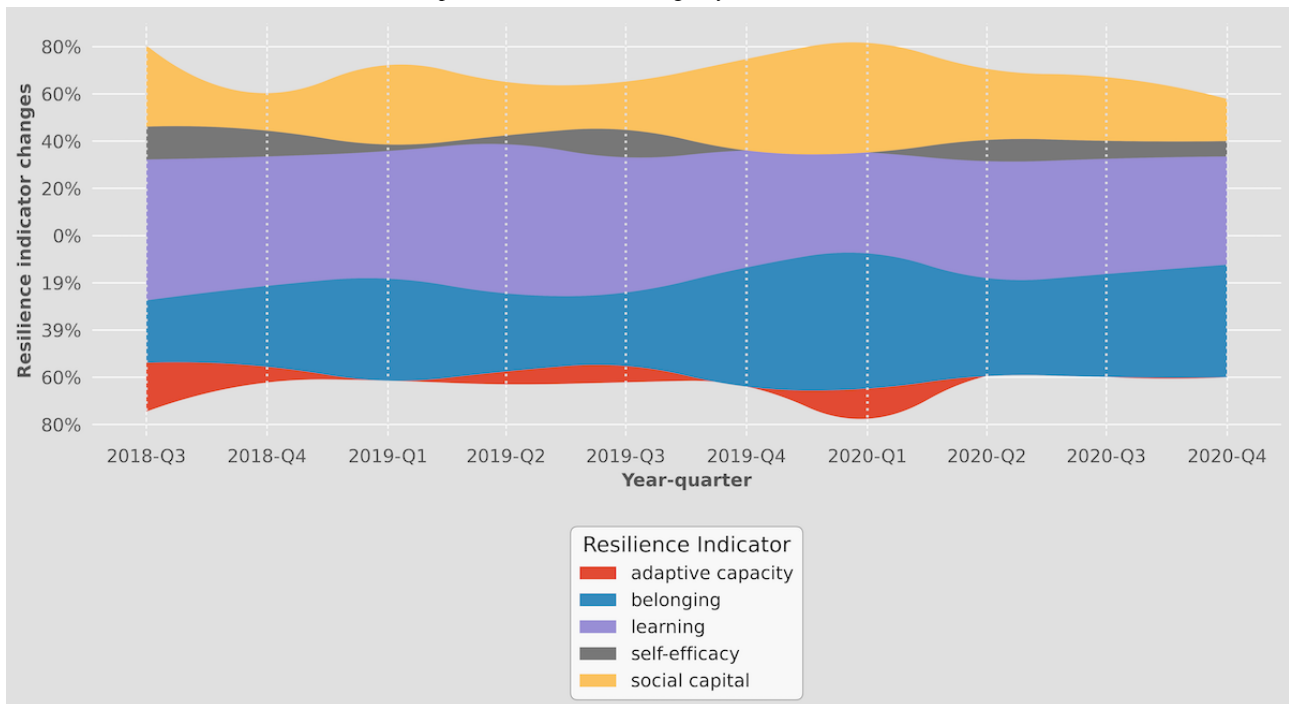
### Results of Resilience Prevalence Analysis

Figure 7 shows the resilience streamgraph that resulted from the resilience prevalence analysis, which identified the dynamics of the realization of resilience indicators over time. This streamgraph was created to visualize our understanding of resilience prevalence [54], which is a popular method for displaying changes in different categories (ie, resilience indicators) over time. Instead of visualizing values as a conventional y-axis, streamgraphs offset the baseline of each “stack” to make it symmetrical around the x-axis. This results in a stream shape that illustrates the change in values over time. Through this streamgraph, we analyzed the patterns of resilience

indicators in terms of their realization over different quarters represented by the x-axis.

A total of 3 findings can be summarized in Figure 7. First, the prevalence of each indicator changes according to the proportion of posts (y-axis), with the height of each individual stream shape showing the prevalence of each indicator over time. The y-axis is not positive or negative but rather shows the best stacking arrangement. Second, “learning” and “belonging” were the top 2 most realized indicators over the entire timeline. The next most common indicator was “social capital,” whereas “self-efficacy” and “adaptive capacity” were less realized during the timeline, with shallow stream shapes appearing and disappearing at different points in time. Third, there were no distinct seasonal peaks and periodic patterns for the top 3 dominant indicators (ie, “learning,” “belonging,” and “social capital”) over the timeline. However, both “belonging” and “social capital” were observed to be slightly stronger at 2020-Q1. By contrast, the realization of both “self-efficacy” and “adaptive capacity” noticeably changed over time: (1) “self-efficacy” was hardly realized at the periods 2019-Q1, 2019-Q2, 2019-Q4, and 2020-Q1, and (2) “adaptive capacity” also seemed to be very weakly realized at the periods 2019-Q1, 2019-Q4, and 2020-Q2 to 2020-Q4. Investigating the possible reasons for these observations may be a useful area for future research.

**Figure 7.** Resilience indicator realization in Schizophrenia: A National Emergency (SANE) Australia.



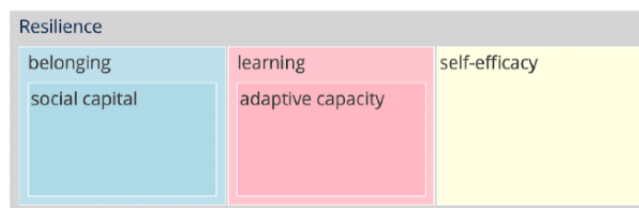
**Phase 4: Resilience Dictionary Building**

**Result of the Resilience Taxonomy Inference**

Given the co-occurrences of the resilience indicators mapped to the posts, we constructed a resilience taxonomy. We used a treemap representation (Figure 8) that shows the corealization of resilience indicators according to 2 levels. The first level has the indicators “belonging” and “learning.” These are higher-level and more general indicators. This means that these indicators were realized at a high prevalence, serving as foundational aspects of resilience. The indicator “social capital” is the more specific indicator subsumed by “belonging.” This means that whenever “social capital” is realized, “belonging” is most likely

to be realized as well. The same association was applied to “learning” and “adaptive capacity.” From another perspective, we can also draw an interpretation relating to the coverage of the 5 resilience indicators, where the coverage of each indicator is indicated by rectangular sizes. “Social capital” is clearly realized in relation to “belonging,” and “adaptive capacity” is realized in relation to “learning.” Colors are used to indicate similar characteristics, where more specific indicators are represented by darker intensities of the same colors. The indicator “self-efficacy” was realized independently, regardless of the other 4 indicators. The emergent taxonomy can be used to gain insights into which indicators are more likely to corealize and what types of conceptual relations are observable between indicators, as inferred by their corealization.

**Figure 8.** The resilience taxonomy induced from the SANE (Schizophrenia: A National Emergency) Australia posts represented as a treemap that shows the coverage of the concept of each resilience indicator by its rectangle size.



**Result of Resilience Dictionary Generation**

Table 3 shows the created resilience dictionary, demonstrating (1) the parent of each resilience indicator represented by “Parent,” (2) topics relevant to each indicator obtained from Table 1 by “Topic,” (3) the relative importance of topics by “Topic Weight,” and 4) the top 10 descriptive terms of each topic by “Topic Words.” The 5 most similar terms for each descriptive term are presented in Multimedia Appendix 1.

In summary, 2 sets of insights were provided by the resilience dictionary. First, sets of topics associated with different

resilience indicators were identified across the entire study period and within each quarter, revealing both constant and periodic topics. These topics and topic words form a useful set of semantic collections that can help characterize each resilience indicator. Second, topic weights, prevalence, and co-occurrences of topics associated with resilience indicators can be used to construct a resilience taxonomy. This taxonomy revealed foundational indicators of resilience—belonging and learning—and more specific and less prevalent indicators—social capital (subsumed by belonging) and adaptive capacity (subsumed by learning)—with self-efficacy realized independently of other indicators.

**Table 3.** The resilience dictionary constructed from this study.

Resilience indicator, parent, and topic	Topic weight (%)	Topic words
<b>Social capital</b>		
<b>Belonging</b>		
Gratitude	3.6	good sound morning news idea hear luck glad wish care
Appreciation	6.1	thank share appreciate reply word kind tag really look ur
Daily routine	7.7	day today photo walk birthday great nice week little yesterday
Loving	3.3	love hug send thinking cat care hon big wish hop
Supporting	2.4	help need support talk really care place try gp reach
Well wishing	1.6	hope ok birthday hey soon happy today enjoy nice glad
Empathizing	1.3	sorry hear really glad know sound care hard moment reach
Fun	1.6	lol ok work yeah today walk week yes maybe home
Friendship	1	friend hello care hug xx send awesome today big mum
<b>Belonging</b>		
Gratitude	3.6	good sound morning news idea hear luck glad wish care
Appreciation	6.1	thank share appreciate reply word kind tag really look ur
Daily routine	7.7	day today photo walk birthday great nice week little yesterday
Loving	3.3	love hug send thinking cat care hon big wish hop
Well wishing	1.6	hope ok birthday hey soon happy today enjoy nice glad
Celebrations	3.1	year happy new christmas birthday merry wish come last family
Empathizing	1.3	sorry hear really glad know sound care hard moment reach
Welcoming	11.5	support welcome member great community mental share health look experience
Fun	1.6	lol ok work yeah today walk week yes maybe home
Friendship	1	friend hello care hug xx send awesome today big mum
<b>Learning</b>		
Reflection	5.2	think people life way make try say need come maybe
Supporting	2.4	help need support talk really care place try gp reach
Knowing	44.8	know say want work people really hard let make life
<b>Adaptive capacity</b>		
<b>Learning</b>		
Reflection	5.2	think people life way make try say need come maybe
<b>Self-efficacy</b>		
Facing difficulties	2.6	feel really feeling make way free bad right moment pain
Sleep	4	sleep night tomorrow last tonight bed today work hon early

## Discussion

### Principal Findings

This paper presents a semiautomatic approach for generating a resilience dictionary from online mental health peer-support forum data. It applies a systematic 4-phase analysis pipeline (Figure 1) that preprocesses raw forum posts, discovers core themes, conceptualizes resilience indicators, and generates a resilience dictionary. We show the promise of exploring how resilience indicators are realized in a web-based mental health community to enrich the characterization of each resilience indicator by showing its range of topics or ways it is discussed

and the interdependencies between different aspects of resilience.

A major contribution of this study is that it provides a replicable method for generating further resilience and other topics or theory-focused dictionaries. The proposed method pipeline can be applied to online support forums hosted by other mental health organizations to better understand their web-based communities and build a deeper understanding of the realization of resilience.

To address RQ1, we present a method for using both the 2-layered NMF topic modeling technique and human knowledge



through thematic analysis to discover and label the 15 substantive topics and map them to the 5 resilience indicators. Each resilience indicator was characterized by its relevant topics (Tables 2 and 3). The analysis showed that we discovered topics related to establishing and maintaining social connections and bonding, building trust, and belonging over time, including through the sharing of milestone events such as Christmas, New Year's Eve, and birthdays, as well as more routine moments. Reflection and knowledge sharing illustrate the importance of learning. In addition, sharing difficulties and ongoing issues associated with mental ill health such as lack of sleep illustrates the role of the forums for participants in navigating self-efficacy and agency amidst the mental health difficulties they face in their daily lives.

We traced the proportion of relevant posts over time to reveal the prevalence of the different aspects of resilience (RQ2). Figure 7 shows the dominance of "learning," "belonging," and "social capital," with "self-efficacy" and "adaptive capacity" appearing less frequently. These findings also help answer RQ3, addressing the relationship between different resilience indicators. To further clarify, we developed a resilience dictionary that shows how each resilience indicator is characterized through a set of relevant topics, the most relevant descriptive terms (including their synonyms) for each topic, and the semantic relationships between resilience indicators observed from the input post data (Table 3 and Multimedia Appendix 1). A taxonomy was developed to establish the relationships of interdependence between different resilience indicators (Figure 8), emphasizing the relationship between generators (belonging and learning) and outcome-oriented (social capital, adaptive capacity, and self-efficacy) indicators of resilience, in line with existing resilience research [27].

A growing body of research has sought to assess the impact of web-based mental health peer support. Peer-support forums offer a form of health intervention that can provide support at scale; however, they also present challenges regarding their design to optimize benefits and their evaluation. New computational methods for NLP, combined with qualitative analysis, provide opportunities to address these challenges. We leveraged NLP-based statistical analysis and qualitative analyses to generate new insights. This work extends beyond the capacity of qualitative content analysis studies that have attempted to identify the effects on resilience through surveys or interviews, cross-sectionally and at a small scale [9,14,18,19,21].

Our work identifies strengths-based indicators of resilience and the relationships between them. This represents an advance or change over previous studies that tended to focus on risk and negative issues regarding mental health, including identifying negative symptoms [31,32], emotions [33,34], and potential risks of self-harm [23]. Although responding to risk and intervening to support people is important, so is the capacity to look at the strengths and resources generated through web-based forums. Complementing these approaches, our work takes a strengths-based approach, with the resultant resilience dictionary able to add a positive context to deepen the understanding of the forum impact and its most effective attributes.

The benefits of the resilience dictionary are that it can be used for (1) showing that resilience, as understood by the current theory, is realized through activity on forums; (2) improving understanding of the aspects of resilience realized on web-based forums; (3) building evidence-based resources for training staff and volunteers by showing how the different aspects of resilience are characterized and how resilience topics are embedded in certain discussions; (4) improving the design and automated moderation of mental health forums by enabling analysis to move beyond risk and harm to strengths-based indicators of resilience; and (5) aiding further text analysis as a vocabulary set for building resilience detection systems that can help mental health care providers to design to improve the outcomes of their services and illustrate their impact.

This study is significant for research and practice in potentially far-reaching areas. First, resilience itself, how it forms, and through what processes and means is poorly understood. Resilience tends to be an ill-defined concept in mental health practice. Here, we have shown how indicators, as part of a conceptual frame, resonate with activity and discussion over time in a web-based forum. Thus, by applying contemporary data analysis techniques to novel data, we have contributed to social and psychological knowledge about this contested concept. Second, for practice, this work shows the potential of (re)using novel data generated through a service to provide useful information that can help design and maintain those services, respond better to health consumers, and show their impact. These are all important issues for nonprofit organizations that struggle to secure and allocate scarce resources. This work should prompt mental health services to gear up their capacity to work with data as it shows considerable potential and innovative uses for data and data skills.

### Limitations

This study has some limitations that could be addressed in future research. First, it focused on creating a resilience dictionary from a single forum and data source (ie, SANE Australia). It would be useful to build resilience dictionaries from multiple data sources and explore similarities and differences. In doing so, we might gain insights into common and distinctive resilience indicators across forums. If resilience realization is different, this could be explored in relation to varying forums and organizational aims, management strategies, and user demographics. Second, the study is "static" in the sense that it used a large existing corpus of historical data. Considering that web-based mental health forums are a growing type of service, it could be useful to investigate how to enable a resilience dictionary to incrementally evolve in response to new forum data.

### Conclusions

In this paper, we presented a promising approach to creating a resilience dictionary that provides insights into the dominant topics on mental health peer-support forums and the way these topics realize different indicators of resilience. The developed mixed NLP and qualitative methods forge new grounds in helping forum providers analyze and understand the impact that the forums provide through a strengths-based analysis of resilience. Along with illustrating the prevalence of different

resilience indicators over time, a taxonomy demonstrates the interdependence of different indicators, revealing which are foundational (belonging and learning) in relation to others (social capital and adaptive capacity). The resulting resilience dictionary offers a benchmark and vocabulary set that can aid

further research. It can also be used to inform automated systems normally predicated on terms associated with risk, harm, and diagnostic indicators of mental ill health or distress, adding a strengths-based approach to the moderation of forum content and interactions.

---

## Acknowledgments

This work was funded by the Australian Research Council Centre of Excellence for Automated Decision Making and Society (CE200100005, 2020-2027), and the Australian Research Council Discovery Project (DP200100419, 2020-2023). The authors would also like to thank Dr Sherridan Emery, University of Tasmania, for her assistance with the topic labeling.

---

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Full resilience dictionary.

[PDF File (Adobe PDF File), 91 KB - formative\_v6i9e39013\_app1.pdf ]

---

## References

1. Mental health: strengthening our response. World Health Organization. 2022 Jun 17. URL: <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response> [accessed 2022-06-17]
2. Anderson K, Priebe S. Concepts of resilience in adolescent mental health research. *J Adolesc Health* 2021 Nov;69(5):689-695. [doi: [10.1016/j.jadohealth.2021.03.035](https://doi.org/10.1016/j.jadohealth.2021.03.035)] [Medline: [34045094](https://pubmed.ncbi.nlm.nih.gov/34045094/)]
3. Moßburger L, Wende F, Brinkmann K, Schmidt T. Exploring online depression forums via text mining: a comparison of Reddit and a curated online forum. In: Proceedings of the 5th Social Media Mining for Health Applications Workshop & Shared Task. 2020 Presented at: SMM4H '20; June 28, 2020; Barcelona, Spain p. 70-81 URL: <https://aclanthology.org/2020.smm4h-1.11> [doi: [10.18653/v1/w18-5906](https://doi.org/10.18653/v1/w18-5906)]
4. Park I, Sarnikar S, Cho J. Disentangling the effects of efficacy-facilitating informational support on health resilience in online health communities based on phrase-level text analysis. *Inf Manag* 2020 Dec;57(8):103372 [FREE Full text] [doi: [10.1016/j.im.2020.103372](https://doi.org/10.1016/j.im.2020.103372)]
5. Ritchie H. Global mental health: five key insights which emerge from the data. Our World In Data. 2018 May 16. URL: <https://ourworldindata.org/global-mental-health> [accessed 2021-09-16]
6. Cook L. Mental health in Australia: a quick guide. Parliament of Australia. 2019 Feb 14. URL: [https://www.aph.gov.au/About\\_Parliament/Parliamentary\\_Departments/Parliamentary\\_Library/pubs/rp/rp1819/Quick\\_Guides/MentalHealth](https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/rp1819/Quick_Guides/MentalHealth) [accessed 2021-09-16]
7. National Study of Mental Health and Wellbeing. First insights from the National Study of Mental Health and Wellbeing, 2020-21. Australian Bureau of Statistics. 2021 Aug 12. URL: <https://www.abs.gov.au/articles/first-insights-national-study-mental-health-and-wellbeing-2020-21> [accessed 2022-04-13]
8. Chan AC, Piehler TF, Ho GW. Resilience and mental health during the COVID-19 pandemic: findings from Minnesota and Hong Kong. *J Affect Disord* 2021 Dec 01;295:771-780 [FREE Full text] [doi: [10.1016/j.jad.2021.08.144](https://doi.org/10.1016/j.jad.2021.08.144)] [Medline: [34517251](https://pubmed.ncbi.nlm.nih.gov/34517251/)]
9. Verdolini N, Amoretti S, Montejó L, García-Rizo C, Hogg B, Mezquida G, et al. Resilience and mental health during the COVID-19 pandemic. *J Affect Disord* 2021 Mar 15;283:156-164 [FREE Full text] [doi: [10.1016/j.jad.2021.01.055](https://doi.org/10.1016/j.jad.2021.01.055)] [Medline: [33556749](https://pubmed.ncbi.nlm.nih.gov/33556749/)]
10. Mayo Clinic Staff. Resilience: Build skills to endure hardship. Mayo Clinic. URL: <https://www.mayoclinic.org/tests-procedures/resilience-training/in-depth/resilience/art-20046311> [accessed 2021-09-17]
11. Depression - Beyond Blue Forums. URL: <https://www.beyondblue.org.au/get-support/online-forums/depression> [accessed 2022-06-01]
12. SANE Australia. URL: <https://www.sane.org/> [accessed 2022-06-01]
13. ReachOut. URL: <https://au.reachout.com/> [accessed 2022-09-07]
14. Hechanova MR, Alianan Jr AS, Jocson RM, Villasanta AP. Online resilience support groups during the COVID-19 pandemic: the Philippine experience. *Ment Health Soc Inclusion* 2022;26(1):56-65. [doi: [10.1108/mhsi-06-2021-0038](https://doi.org/10.1108/mhsi-06-2021-0038)]
15. Fisher EB, Miller SM, Evans M, Luu SL, Tang PY, Dreyer Valovcin D, et al. COVID-19, stress, trauma, and peer support-observations from the field. *Transl Behav Med* 2020 Aug 07;10(3):503-505 [FREE Full text] [doi: [10.1093/tbm/ibaa056](https://doi.org/10.1093/tbm/ibaa056)] [Medline: [32569372](https://pubmed.ncbi.nlm.nih.gov/32569372/)]

16. McCosker A. Engaging mental health online: insights from beyondblue's forum influencers. *New Media Soc* 2018 Jul 11;20(12):4748-4764 [FREE Full text] [doi: [10.1177/1461444818784303](https://doi.org/10.1177/1461444818784303)]
17. Gillard S, Foster R, Gibson S, Goldsmith L, Marks J, White S. Describing a principles-based approach to developing and evaluating peer worker roles as peer support moves into mainstream mental health services. *Ment Health Soc Inclusion* 2017 Jun 12;21(3):133-143. [doi: [10.1108/mhsl-03-2017-0016](https://doi.org/10.1108/mhsl-03-2017-0016)]
18. Wu Y, Sang ZQ, Zhang XC, Margraf J. The relationship between resilience and mental health in Chinese college students: a longitudinal cross-lagged analysis. *Front Psychol* 2020 Feb 5;11:108 [FREE Full text] [doi: [10.3389/fpsyg.2020.00108](https://doi.org/10.3389/fpsyg.2020.00108)] [Medline: [32116918](https://pubmed.ncbi.nlm.nih.gov/32116918/)]
19. Riehm KE, Brenneke SG, Adams LB, Gilan D, Lieb K, Kunzler AM, et al. Association between psychological resilience and changes in mental distress during the COVID-19 pandemic. *J Affect Disord* 2021 Mar 01;282:381-385 [FREE Full text] [doi: [10.1016/j.jad.2020.12.071](https://doi.org/10.1016/j.jad.2020.12.071)] [Medline: [33421866](https://pubmed.ncbi.nlm.nih.gov/33421866/)]
20. Building your resilience. American Psychological Association. URL: <https://www.apa.org/topics/resilience> [accessed 2021-09-16]
21. Liu WJ, Zhou L, Wang XQ, Yang BX, Wang Y, Jiang JF. Mediating role of resilience in relationship between negative life events and depression among Chinese adolescents. *Arch Psychiatr Nurs* 2019 Dec;33(6):116-122. [doi: [10.1016/j.apnu.2019.10.004](https://doi.org/10.1016/j.apnu.2019.10.004)] [Medline: [31753216](https://pubmed.ncbi.nlm.nih.gov/31753216/)]
22. Conway M, Hu M, Chapman WW. Recent advances in using natural language processing to address public health research questions using social media and ConsumerGenerated data. *Yearb Med Inform* 2019 Aug;28(1):208-217 [FREE Full text] [doi: [10.1055/s-0039-1677918](https://doi.org/10.1055/s-0039-1677918)] [Medline: [31419834](https://pubmed.ncbi.nlm.nih.gov/31419834/)]
23. Milne DN, McCabe KL, Calvo RA. Improving moderator responsiveness in online peer support through automated triage. *J Med Internet Res* 2019 Apr 26;21(4):e11410 [FREE Full text] [doi: [10.2196/11410](https://doi.org/10.2196/11410)] [Medline: [31025945](https://pubmed.ncbi.nlm.nih.gov/31025945/)]
24. r/depression stats. Reddit. URL: <https://subredditstats.com/r/depression> [accessed 2022-06-01]
25. HealthBoards. URL: <https://www.healthboards.com> [accessed 2022-06-01]
26. Inspire. URL: <https://www.inspire.com> [accessed 2022-06-01]
27. HealthUnlocked. URL: <https://www.healthunlocked.com> [accessed 2022-06-01]
28. Davcheva E. Text mining mental health forums - learning from user experiences. In: Proceedings of the 26th European Conference on Information Systems. 2018 Presented at: ECIS '18; June 23-28, 2018; Portsmouth, UK p. 91 URL: [https://aisel.aisnet.org/ecis2018\\_rp/91/](https://aisel.aisnet.org/ecis2018_rp/91/)
29. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. *JMIR Public Health Surveill* 2020 May 22;6(2):e19447. [doi: [10.2196/19447](https://doi.org/10.2196/19447)] [Medline: [32412418](https://pubmed.ncbi.nlm.nih.gov/32412418/)]
30. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. *Nat Lang Eng* 2017 Jan 30;23(5):649-685. [doi: [10.1017/S1351324916000383](https://doi.org/10.1017/S1351324916000383)]
31. Park A, Conway M. Longitudinal changes in psychological states in online health community members: understanding the long-term effects of participating in an online depression community. *J Med Internet Res* 2017 Mar 20;19(3):e71 [FREE Full text] [doi: [10.2196/jmir.6826](https://doi.org/10.2196/jmir.6826)] [Medline: [28320692](https://pubmed.ncbi.nlm.nih.gov/28320692/)]
32. Yazdavar AH, Mahdavinejad MS, Bajaj G, Romine W, Sheth A, Monadjemi AH, et al. Multimodal mental health analysis in social media. *PLoS One* 2020 Apr 10;15(4):e0226248 [FREE Full text] [doi: [10.1371/journal.pone.0226248](https://doi.org/10.1371/journal.pone.0226248)] [Medline: [32275658](https://pubmed.ncbi.nlm.nih.gov/32275658/)]
33. Patton DU, MacBeth J, Schoenebeck S, Shear K, McKeown K. Accommodating grief on Twitter: an analysis of expressions of grief among gang involved youth on Twitter using qualitative analysis and natural language processing. *Biomed Inform Insights* 2018 Apr 3;10:1178222618763155 [FREE Full text] [doi: [10.1177/1178222618763155](https://doi.org/10.1177/1178222618763155)] [Medline: [29636619](https://pubmed.ncbi.nlm.nih.gov/29636619/)]
34. Loveys K, Crutchley P, Wyatt E, Coppersmith G. Small but mighty: affective micropatterns for quantifying mental health from social media language. In: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology. 2017 Presented at: CLPsych '17; August 3, 2017; Vancouver, Canada p. 85-95. [doi: [10.18653/v1/w17-3110](https://doi.org/10.18653/v1/w17-3110)]
35. SANE Australia Annual Highlights 2019-2020. SANE Australia. 2021. URL: <https://www.sane.org/about-sane/annual-reports-highlights> [accessed 2022-06-01]
36. Berkes F, Ross H. Community resilience: toward an integrated approach. *Soc Natural Resources* 2013 Jan;26(1):5-20 [FREE Full text] [doi: [10.1080/08941920.2012.736605](https://doi.org/10.1080/08941920.2012.736605)]
37. Brown K, Westaway E. Agency, capacity, and resilience to environmental change: lessons from human development, well-being, and disasters. *Annu Rev Environ Resour* 2011 Nov 21;36(1):321-342 [FREE Full text] [doi: [10.1146/annurev-environ-052610-092905](https://doi.org/10.1146/annurev-environ-052610-092905)]
38. Skerratt S. Enhancing the analysis of rural community resilience: evidence from community land ownership. *J Rural Stud* 2013 Jul;31:36-46 [FREE Full text] [doi: [10.1016/j.jrurstud.2013.02.003](https://doi.org/10.1016/j.jrurstud.2013.02.003)]
39. Fazey I, Fazey JA, Fischer J, Sherren K, Warren J, Noss RF, et al. Adaptive capacity and learning to learn as leverage for social-ecological resilience. *Front Ecol Environ* 2007 Sep;5(7):375-380 [FREE Full text] [doi: [10.1890/1540-9295\(2007\)5\[375:acaltl\]2.0.co;2](https://doi.org/10.1890/1540-9295(2007)5[375:acaltl]2.0.co;2)]
40. Reisenbichler M, Reutterer T. Topic modeling in marketing: recent advances and research opportunities. *J Bus Econ* 2019;89(3):327-356. [doi: [10.1007/s11573-018-0915-7](https://doi.org/10.1007/s11573-018-0915-7)]

41. Albalawi R, Yeap TH, Benyoucef M. Using topic modeling methods for short-text data: a comparative analysis. *Front Artif Intell* 2020 Jul 14;3:42 [FREE Full text] [doi: [10.3389/frai.2020.00042](https://doi.org/10.3389/frai.2020.00042)] [Medline: [33733159](https://pubmed.ncbi.nlm.nih.gov/33733159/)]
42. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature* 1999 Oct 21;401(6755):788-791. [doi: [10.1038/44565](https://doi.org/10.1038/44565)] [Medline: [10548103](https://pubmed.ncbi.nlm.nih.gov/10548103/)]
43. O'Callaghan D, Greene D, Carthy J, Cunningham P. An analysis of the coherence of descriptors in topic modeling. *Expert Syst Appl* 2015 Aug;42(13):5645-5657 [FREE Full text] [doi: [10.1016/j.eswa.2015.02.055](https://doi.org/10.1016/j.eswa.2015.02.055)]
44. Aizawa A. An information-theoretic perspective of tf-idf measures. *Inf Process Manag* 2003 Jan;39(1):45-65 [FREE Full text] [doi: [10.1016/s0306-4573\(02\)00021-3](https://doi.org/10.1016/s0306-4573(02)00021-3)]
45. Belford M, Mac Namee B, Greene D. Stability of topic modeling via matrix factorization. *Expert Syst Appl* 2018 Jan;91(C):159-169. [doi: [10.1016/j.eswa.2017.08.047](https://doi.org/10.1016/j.eswa.2017.08.047)]
46. Isoaho K, Gritsenko D, Mäkelä E. Topic modeling and text analysis for qualitative policy research. *Policy Stud J* 2021 Feb;49(1):300-324 [FREE Full text] [doi: [10.1111/psj.12343](https://doi.org/10.1111/psj.12343)]
47. Kang YB, Haghigh PD, Burstein F. TaxoFinder: a graph-based approach for taxonomy learning. *IEEE Trans Knowl Data Eng* 2016 Feb 1;28(2):524-536. [doi: [10.1109/tkde.2015.2475759](https://doi.org/10.1109/tkde.2015.2475759)]
48. Kang YB, Woo J, Kneebone L, Sellis T. Methodology for refining subject terms and supporting subject indexing with taxonomy: a case study of the APO digital repository. *Decis Support Syst* 2021 Jul;146:113542 [FREE Full text] [doi: [10.1016/j.dss.2021.113542](https://doi.org/10.1016/j.dss.2021.113542)]
49. de Knijff J, Frasinca F, Hogenboom F. Domain taxonomy learning from text: the subsumption method versus hierarchical clustering. *Data Knowl Eng* 2013 Jan;83:54-69 [FREE Full text] [doi: [10.1016/j.datak.2012.10.002](https://doi.org/10.1016/j.datak.2012.10.002)]
50. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *Comput Res Repository* 2013 [FREE Full text]
51. Chen Y, Zhang H, Liu R, Ye Z, Lin J. Experimental explorations on short text topic mining between LDA and NMF based schemes. *Knowl Based Syst* 2019 Jan;163:1-13 [FREE Full text] [doi: [10.1016/j.knosys.2018.08.011](https://doi.org/10.1016/j.knosys.2018.08.011)]
52. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
53. Miller V. New media, networking and phatic culture. *Convergence* 2008 Nov 01;14(4):387-400. [doi: [10.1177/1354856508094659](https://doi.org/10.1177/1354856508094659)]
54. Bider L. Streamgraphs: how to make them and what you need to know. *Flourish*. 2020 Oct 19. URL: <https://flourish.studio/2020/10/19/streamgraphs/> [accessed 2021-10-10]

## Abbreviations

- NLP:** natural language processing  
**NMF:** nonnegative matrix factorization  
**RQ:** research question  
**SANE:** Schizophrenia: A National Emergency

*Edited by T Leung; submitted 25.04.22; peer-reviewed by S Chen, G Makwana; comments to author 19.05.22; revised version received 06.06.22; accepted 18.08.22; published 22.09.22.*

*Please cite as:*

*Kang YB, McCosker A, Kamstra P, Farmer J*

*Resilience in Web-Based Mental Health Communities: Building a Resilience Dictionary With Semiautomatic Text Analysis*

*JMIR Form Res* 2022;6(9):e39013

URL: <https://formative.jmir.org/2022/9/e39013>

doi: [10.2196/39013](https://doi.org/10.2196/39013)

PMID: [36136394](https://pubmed.ncbi.nlm.nih.gov/36136394/)

©Yong-Bin Kang, Anthony McCosker, Peter Kamstra, Jane Farmer. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 22.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Relative Effectiveness of Social Media, Dating Apps, and Information Search Sites in Promoting HIV Self-testing: Observational Cohort Study

Chrysovalantis Stafylis<sup>1</sup>, MD, MPH; Gabriella Vavala<sup>2</sup>, BA; Qiao Wang<sup>2</sup>, MPH; Bethany McLeman<sup>3</sup>, BA; Shea M Lemley<sup>3</sup>, PhD; Sean D Young<sup>4,5</sup>, PhD; Haiyi Xie<sup>3</sup>, PhD; Abigail G Matthews<sup>6</sup>, PhD; Neal Oden<sup>6</sup>, PhD; Leslie Revoredó<sup>3</sup>, PhD; Dikla Shmueli-Blumberg<sup>6</sup>, PhD; Emily G Hichborn<sup>3</sup>, BSc; Erin McKelle<sup>7</sup>, MA; Landhing M Moran<sup>3</sup>, PhD; Petra Jacobs<sup>8</sup>, MD, MHS; Lisa A Marsch<sup>3</sup>, PhD; Jeffrey D Klausner<sup>1</sup>, MD, MPH

<sup>1</sup>Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA, United States

<sup>2</sup>Division of Infectious Diseases, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States

<sup>3</sup>Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States

<sup>4</sup>Department of Emergency Medicine, School of Medicine, University of California, Irvine, Irvine, CA, United States

<sup>5</sup>Department of Informatics, Bren School of Information and Computer Sciences, University of California, Irvine, Irvine, CA, United States

<sup>6</sup>The Emmes Company LLC, Rockville, MD, United States

<sup>7</sup>Education, Training and Research Associates, Oakland, CA, United States

<sup>8</sup>National Institute on Aging, Bethesda, MD, United States

**Corresponding Author:**

Chrysovalantis Stafylis, MD, MPH  
Department of Population and Public Health Sciences  
University of Southern California  
2001 N. Soto St.  
Los Angeles, CA, 90033  
United States  
Phone: 1 626 757 8602  
Email: [chrstafylis@outlook.com](mailto:chrstafylis@outlook.com)

## Abstract

**Background:** Social media sites, dating apps, and information search sites have been used to reach individuals at high risk for HIV infection. However, it is not clear which platform is the most efficient in promoting home HIV self-testing, given that the users of various platforms may have different characteristics that impact their readiness for HIV testing.

**Objective:** This study aimed to compare the relative effectiveness of social media sites, dating apps, and information search sites in promoting HIV self-testing among minority men who have sex with men (MSM) at an increased risk of HIV infection. Test kit order rates were used as a proxy to evaluate promotion effectiveness. In addition, we assessed differences in characteristics between participants who ordered and did not order an HIV test kit.

**Methods:** Culturally appropriate advertisements were placed on popular sites of three different platforms: social media sites (Facebook, Instagram), dating apps (Grindr, Jack'D), and information search sites (Google, Bing). Advertisements targeted young (18-30 years old) and minority (Black or Latinx) MSM at risk of HIV exposure. Recruitment occurred in 2 waves, with each wave running advertisements on 1 platform of each type over the same period. Participants completed a baseline survey assessing sexual or injection use behavior, substance use including alcohol, psychological readiness to test, attitudes toward HIV testing and treatment, and HIV-related stigma. Participants received an electronic code to order a free home-based HIV self-test kit. Follow-up assessments were conducted to assess HIV self-test kit use and uptake of pre-exposure prophylaxis (PrEP) at 14 and 60 days post enrollment.

**Results:** In total, 271 participants were enrolled, and 254 were included in the final analysis. Among these 254 participants, 177 (69.7%) ordered a home HIV self-test kit. Most of the self-test kits were ordered by participants enrolled from dating apps. Due to waves with low enrollment, between wave statistical comparisons were not feasible. Within wave comparison revealed that Jack'D showed higher order rates (3.29 kits/day) compared to Instagram (0.34 kits/day) and Bing (0 kits/day). There were

no associations among self-test kit ordering and HIV-related stigma, perceptions about HIV testing and treatment, and mistrust of medical organizations.

**Conclusions:** Our findings show that using popular dating apps might be an efficient way to promote HIV self-testing. Stigma, perceptions about HIV testing and treatment, or mistrust of medical organizations may not affect order rates of HIV test kits promoted on the internet.

**Trial Registration:** ClinicalTrials.gov NCT04155502; <https://clinicaltrials.gov/ct2/show/NCT04155502>

**International Registered Report Identifier (IRRID):** RR2-10.2196/20417

(*JMIR Form Res* 2022;6(9):e35648) doi:[10.2196/35648](https://doi.org/10.2196/35648)

## KEYWORDS

HIV prevention; PrEP; home HIV test; social media; dating apps; search engines; HIV; human immunodeficiency virus; self-testing; infection; digital health; health promotion; MSM; pre-exposure prophylaxis; medical information

## Introduction

The incidence of HIV infection remains high among minority men who have sex with men (MSM) [1]. Frequent testing for HIV infection can identify new infections early, and it is essential in ending the HIV epidemic [2]. HIV self-testing is an alternative HIV screening method that is commercially available, approved by the Food and Drug Administration, and can reach individuals who have never tested before. It can reach populations at risk, such as Black and Latinx individuals, identify new cases of HIV infection [3-6], and lead individuals to seek additional HIV prevention options, such as testing for sexually transmitted infections or pre-exposure prophylaxis (PrEP) [7]. Prevention studies and public health programs have been adopting HIV self-tests [8,9] and combining them with new technologies, such as smartphone apps [10] or smart devices [4], to reach populations with high incidence of HIV infection, such as Black and Latinx MSM. Despite multiple efforts, the uptake of HIV testing remains inadequate, especially among individuals at high risk for HIV infection [11]. Thus, optimizing the promotion of HIV testing is important.

Due to their extensive popularity, social media sites and dating apps have been used to promote and recruit participants for HIV prevention research studies with high rates of success [5,12-14]. According to a recent Centers for Disease Control and Prevention (CDC) report reviewing HIV self-testing programs, 27 health departments and community organizations [9] used multiple platforms for promotion, mainly social media (19/27) followed by “traditional” printed media (9/27) and dating apps (6/27). Compared to in-person recruitment, web-based platforms have the capacity to reach a high number of difficult-to-reach populations and individuals at risk [5,14,15], overcoming stigma or other logistic obstacles [15,16] in a cost-efficient manner [16,17]. Indeed, the New York Department of Public Health used advertisements on social media, dating apps, and websites to reach 28,921 users, identifying 17,383 eligible MSM, transgender, and gender nonconforming individuals during its HIV self-testing campaign. Most of the participants were under the age of 35 years and identified as Black or Latinx. In addition, the first wave of this campaign reached 3359 users in only 23 days, distributing 2497 home test kit voucher codes to eligible users [18]. Social media and dating apps have been widely adopted as means of promoting HIV home testing. Although different from dating apps and social media sites, information

search sites (eg, Google) are commonly used for seeking information on HIV testing and PrEP [19,20] and could represent a promising outreach avenue. Their use for enrollment and HIV testing promotion has not been evaluated.

However, little is known about the relative effectiveness of these different web-based platforms (ie, social media, dating apps, and information search sites) in promoting HIV self-testing. Parker et al [21] conducted a secondary analysis in a study enrolling substance-using sexual and gender minority adolescents and young adults to evaluate the efficacy of their enrollment strategy. The study used multiple methods to enroll participants, including social media platforms (Facebook, Instagram, Twitter, Tumblr), dating apps (Grindr, Scruff, Jack'D), internet-based health boards, and venue-based enrollment. They recorded 17,328 visits to the eligibility screener on the landing page, with a 36.2% (6274/17,328) screener survey completion ratio. Researchers identified 580 participants among those who consented and were eligible to participate (580/623, 93.1%), indicating a high recruitment proportion. The majority of their participants were enrolled from Facebook, Instagram, and Grindr. Studies and programs use these platforms based on the experience of previous studies and expert recommendations [22].

Data on the effectiveness of public health promotion through different platforms leading to testing or PrEP are missing. We can only infer the effectiveness of promotion indirectly, as head-to-head comparisons of the effectiveness of the different platforms and sites to reach individuals for public health promotion are missing. This would allow researchers and prevention programs to optimize their budget and strategy. The primary objective of this study was to compare ordering of HIV self-testing kits among users recruited through 3 different types of web-based platforms, including social media, dating apps, and information search sites. Test kit ordering was used as a proxy for analyzing the effectiveness of promoting HIV self-testing on different sites. The secondary goal was to evaluate the association of key moderating variables—substance use, psychological readiness to test, and perceptions and attitudes related to HIV testing—with the ordering of HIV self-testing kits.

## Methods

### Recruitment

In this longitudinal observational cohort study, advertisements promoting free HIV self-testing were placed on three platform types: social media (Facebook, Instagram), dating apps (Grindr, Hornet), and information search sites (Google, Bing) (Table S1 in [Multimedia Appendix 1](#)). The advertisements were organized in 2 “waves,” with each wave consisting of 1 social media website, 1 dating app, and 1 information search site. The Wave 1 (Facebook, Grindr, Google) recruitment stopped early, as Grindr unexpectedly stopped running all self-service platform advertisements (including the study advertisement) due to a change in corporate ownership [23,24]. We continued with Wave 2 (Instagram, Jack’D, Bing) as planned and relaunched Wave 1 (Facebook, Grindr, Google) once Grindr access was restored.

Before launching each wave, we allocated the same amount of funds for each of the 3 sites and optimized them to run for at least 30 calendar days by dividing the available funds in the prespecified promotional period. However, due to slow enrollment during the COVID-19 pandemic, we extended the second phase of Wave 1 up to 63 days. The advertisement used on social media and dating apps was an image that included a person and text (“Get a FREE HIV test”), whereas promotional keywords related to HIV testing and PrEP were selected for information search sites (as images are not allowed). The same image and keywords were used in all waves. The advertisements were launched in the District of Columbia (DC) and 8 states (Florida, Georgia, Louisiana, Maryland, Mississippi, Nevada, South Carolina, and Texas), which were selected based on their high incidence of HIV infection. More information regarding the promotional campaign can be found in the published protocol [25].

Upon clicking on the study advertisement, website users landed on the study information page, where they received general information about the study, underwent eligibility screening, and reviewed study procedures. Following electronic informed consent, participants completed the baseline assessment and were emailed a unique electronic code to order their HIV home self-test kit through Orasure.com (Bethlehem, PA). Participants also received an electronic coupon for a free telemedicine PrEP visit. Participants were followed up at 14 and 60 days after enrollment. At follow-up, participants were asked about their HIV self-test use and self-test results; depending on their self-test result, they were asked if they visited a PrEP provider and started PrEP, as well as their opinions on PrEP. If they tested positive for HIV antibodies with the home self-test kit, they were asked if they had visited a clinic for confirmatory testing and HIV treatment. In addition, we tracked test kit orders through automated reports, collected anonymous advertisement metrics through the web applications of the platforms, and recorded the costs for each promotion site and wave.

### Inclusion Criteria

We enrolled MSM aged 18-30 years who identified as Latinx or Black/African American people (including multiracial and multiethnic individuals of these groups); they reported having

condomless anal sex in the past 90 days or having more than 1 male sex partner in the past 90 days. Participants were considered ineligible if they were HIV-positive, if they were tested for HIV infection in the past 90 days, and if they were taking PrEP currently or at any time during the past 6 months before enrollment.

### Outcome

The primary outcome was the number of HIV self-test kits ordered per day through each type of internet-based platform (social media, dating, information site) during the period in which each wave was operational. As a secondary outcome, we explored the association of reported substance use, stage of change for HIV testing based on the transtheoretical model, attitudes toward HIV testing and treatment, HIV-related stigma, medical mistrust, and opinions about PrEP measures and self-test kit ordering. As an exploratory outcome, we recorded the advertisement metrics of each campaign to measure differences in the reach and cost.

### Assessments

Study assessments are described in the protocol [25]. Participants were asked to self-report test kit and PrEP use. We calculated the substance-specific TAPS (Tobacco, Alcohol, Prescription medication, and other Substance) tool score [26] of each participant. For each substance, a score of 1 was classified as “problem use” (low-risk substance use), whereas a score of 2 or higher was classified as “high-risk substance use.” We collected participants’ opinions about HIV treatment using a 10-item questionnaire [27]. Each question was presented as a visual analog scale (eg, slider) with “strongly disagree” and “strongly agree” anchoring the 2 extremes. We assumed an underlying continuous, linear relationship between the 2 anchors, and data for opinions about HIV treatment are presented as the mean score for each question with its SD. PrEP opinions, barriers, and facilitators were collected using a 5-point Likert scale (ranging from “Not at all important” to “Extremely important”).

Finally, we monitored the performance of the promotional campaigns using the impressions (number of times the advertisement is shown on a screen), clicks (number of times the advertisement was “clicked”), click-through rate (clicks/impressions), and funds spent.

### Statistical Analysis

Participants who were enrolled from Google and Facebook while Grindr was inactive (early during Wave 1) were excluded from analyses. This ensured that we included data when all 3 sites were active and thus had an equal chance to enroll participants. Participants who did not order a test kit within 60 days of the test code being emailed to them were classified as “not ordered a self-test kit.” The 2 advertisement periods of Wave 1 were combined before analysis. Prior to statistical modeling, the number of HIV home self-test kits ordered from each platform, specific platform types (sites), and number days of recruitment in each wave were summarized. In addition, the observed daily self-test kit order rates for each site and platform type were calculated (order rate = number of orders / number of advertising days during each wave).

Per our primary research question, we intended to determine the statistical difference in the self-test kit ordering rates by platform type (social media, information search site, and dating app) using a Poisson regression model; however, due to significant platform-by-wave interactions and widely differing order rates between sites within the same platform, it was not appropriate to combine or pool sites across the same platform for statistical evaluation of the platform difference. Therefore, we compared the specific platform differences in terms of the order rates within the same wave. We conducted pairwise comparison for all 6 sites from the 2 waves with multiple testing adjustments using the Hochberg method [28].

Demographic and baseline characteristics were presented using summary statistics. Continuous variables were summarized using percentiles (median, and 25th and 75th percentiles), and means with their SDs. Categorical variables were summarized with frequencies and percentages. To assess differences in the measures between participants who ordered a test kit and those who did not order a test kit, we used the Student *t* test for continuous variables, Fisher exact test for categorical variables, and Wilcoxon rank sum test for Likert responses. Data analysis was carried out using Statistical Analysis Software (version 9.4, SAS Institute).

### Adaptations Due to COVID-19 Pandemic

We conducted a third wave of promotion and enrollment on Twitter, Yahoo, and Hornet. This wave was conducted between April 6, 2020, and May 6, 2020, during the first days of the public health emergency proclamation. Despite the promotional waves being active, no participants were enrolled, and no test kits were ordered during Wave 3, which made our statistical model inestimable. As enrollment during this period does not reflect “expected conditions” and scientific comparisons would not be accurate, we decided to exclude Wave 3 from all the analyses.

### Sensitivity Analyses

We conducted 3 sensitivity analyses using the statistical approach, Poisson regression, and posthoc contrast. The primary sensitivity analysis included any self-test kits ordered at any time during the study (ie, outside of the 60-day window for the primary analysis) and by any participants in the validated participant population. The second sensitivity analysis attempted to address the fact that Wave 1 occurred in 2 phases because 1 promotional platform (Grindr) stopped all advertising. The final sensitivity analysis assessed the impact of the COVID-19 pandemic.

### Missing Data

The analysis of primary outcomes does not include missing data. In questions where participants could “skip” and not respond, the answer was classified as “missing” and was not included in the calculation of those variable frequencies.

### Ethics Approval

This study (trial registration: NCT04155502) was reviewed and approved by the Institutional Review Board at the University of California, Los Angeles (IRB #18-001580).

## Results

### Baseline Characteristics

Between January and September 2020, a total of 10,669 individuals visited the study website, directed from study advertisements placed on the platform sites across all waves. During the study period, 254 participants were enrolled from 6 platform sites. The majority were enrolled from urban areas of Texas, Florida, DC, and Georgia. The average age (SD) of participants was 24.4 years (SD 3.7 years). Most (199/254, 78.4%) participants identified as Black/African American and 26% (66/254) reported that they were Latinx.

The median number of sex partners in the past 90 days was 4 (IQR 3-6). Among the 254 participants, 210 (82.7%) participants reported receptive condomless anal sex during the past 90 days. Only 23 (8.9%) participants received PrEP before. When asked about condom use, 5 (2%) reported that they always used a condom, whereas 36 (14.2%) said that they never used condoms. Most of the participants (191, 75.2%) tested for HIV infection in the past. Among those tested in the past, the median (IQR) time since their last test was 11 months (6-21). Participants who never previously tested reported that their main reasons for not testing were their fear of obtaining a positive HIV result and their belief that HIV exposure was unlikely. Table 1 presents the participant demographics and behaviors.

In terms of HIV home test kit use, 131 out of the 177 participants (74%) reported a self-test result, with 11 of the 131 participants (8.4%) reporting a positive HIV test result; 9 of these 11 (82%) reported that they sought confirmatory testing and 4 of these 9 (44.4%) had started treatment for HIV. Among the 120 participants who reported a negative test result for HIV infection, 13 (11%) reported visiting a provider to discuss PrEP or reported starting PrEP (Figure 1).

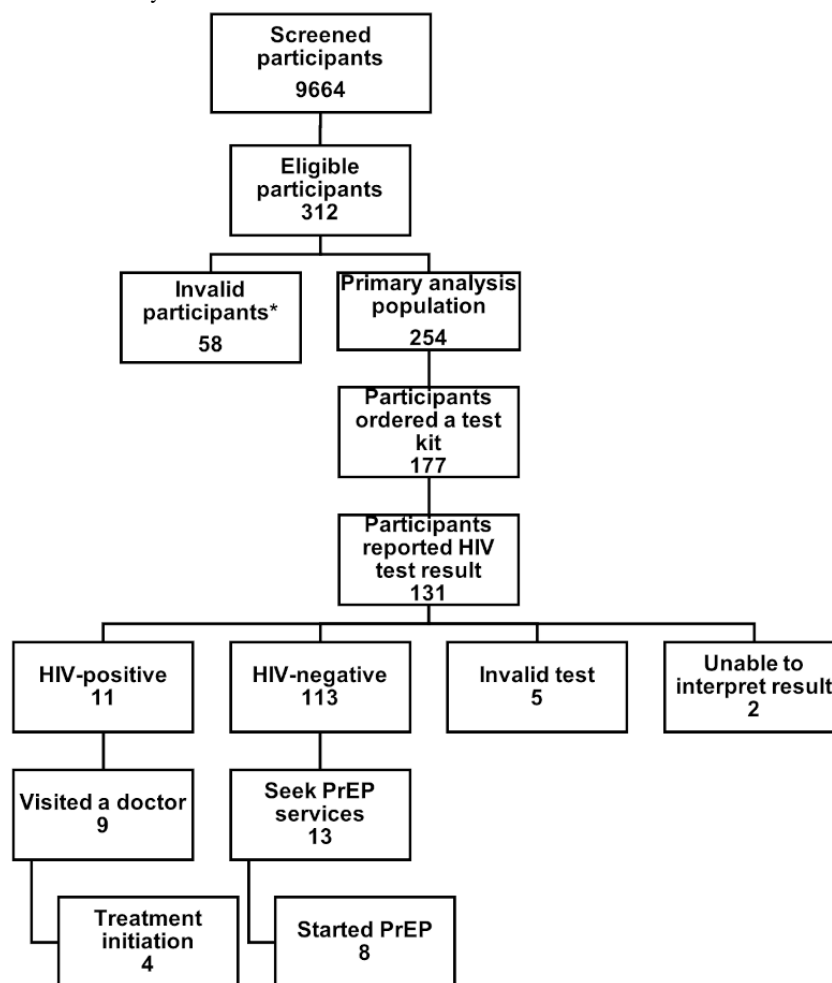


**Table 1.** Summary of National Institute on Drug Abuse Clinical Trials Network Social Media Pre-exposure Prophylaxis Study, 2020, population sociodemographic and behavioral characteristics (N=254).

Characteristic	Value
Age in years, median (IQR)	25 (23-27)
<b>Ethnicity, n (%)</b>	
Hispanic/Latinx	66 (26)
<b>Race, n (%)</b>	
American Indian or Alaskan Native	1 (0.4)
Black or African American	196 (78.4)
White	28 (11.2)
Other	14 (5.6)
Multiracial	11 (4.4)
<b>History of PrEP<sup>a</sup> uptake, n (%)</b>	
Never taken PrEP	232 (91.3)
In the past 6 months	22 (8.9)
Number of male sex partners in the past 90 days, median (IQR)	4 (3-6)
<b>Condom use, n (%)</b>	
Never	36 (14.2)
Sometimes	108 (42.5)
About half the time	37 (14.5)
Most of the time	68 (26.8)
Always	5 (2)
Condomless receptive anal sex in the past 90 days, n (%)	210 (82.7)
Ever tested for HIV during lifetime, n (%)	191 (75.2)
<b>If tested for HIV, median (IQR)</b>	
Months since last HIV test	11 (6-21)
If not tested for HIV, n (%)	63 (24.8%)
<b>Main reasons cited by the 63 participants for not getting tested, n (%)</b>	
Unlikely to be exposed to HIV	8 (12.7)
Afraid of testing HIV-positive	26 (41.3)
Did not want to think about HIV/HIV-positive	8 (12.7)
Worried about names being reported if positive	3 (4.8)
Dislike for needles	5 (8)
Unable to trust that the results will be confidential	3 (4.8)
Unaware of where to get tested	7 (11.1)
Other reasons	3 (4.8)

<sup>a</sup>PrEP: pre-exposure prophylaxis.

**Figure 1.** Enrollment, HIV home test kit use, pre-exposure prophylaxis uptake, and linkage to care among participants of the National Institute on Drug Abuse Clinical Trials Network Social Media PrEP Study, 2020. PrEP: pre-exposure prophylaxis. \*Invalid participants include duplicate entries, fake accounts, and participants outside the country.



### Primary Outcome

[Table 2](#) summarizes the analysis results for the primary outcome. In total, 177 of the 254 participants ordered test kits during the study period. Overall, those recruited through dating apps had the highest order rate (1.24 kits/day), followed by social media platforms (0.24 kits/day) and information search platforms (0.16 kits/day; [Table 2](#)). Pairwise contrasts between the platforms showed that in Wave 1, there was no statistically significant difference across the specific platforms. Specifically, the Hockberg-adjusted  $P=.59$  for all the following pairwise contrasts in Wave 1: Facebook (social media) vs Google (information site), Facebook (social media) vs Grinder (dating app), and

Google vs Grinder contrasts (Note: False discovery-adjusted  $P$  values were different between each pair contrast but not significant.). However, in Wave 2, there was a statistically significant difference across the platforms with Jack'D (dating app) being the most effective site (3.29 kits/day), compared to Instagram (0.34 kits/day) and Bing (0 kits/day). Specifically, the Hockberg-adjusted  $P=.002$  for Bing (information site) vs Instagram (social media) contrast;  $P<.001$  for the Bing (Information site) vs Jack'D (dating app) contrast and for the Instagram vs Jack'D contrast. All 3 types of primary outcome sensitivity analyses showed results similar to the primary analysis (see [Multimedia Appendix 2](#) for details).

**Table 2.** Number and rate of HIV home self-test kits ordered through promotional platforms by wave per protocol sample in the National Institute on Drug Abuse Clinical Trials Network Social Media Pre-exposure Prophylaxis Study, 2020 (N=254).

Type of platform	Wave	Number of days for each wave	Number of test kits ordered	Order rate (ordered test kits/day)
<b>Social media site</b>				
Facebook	1 <sup>a</sup>	70	13	0.19
Instagram	2	38	13	0.34
Subtotal	N/A	108	26	0.24
<b>Dating app</b>				
Grindr	1 <sup>a</sup>	70	9	0.13
Jack'D	2	38	125	3.29
Subtotal	N/A	108	134	1.24
<b>Information search site</b>				
Google	1 <sup>a</sup>	70	17	0.24
Bing	2	38	0	0.00
Subtotal	N/A	108	17	0.16
Total	N/A	108	177	1.64

<sup>a</sup>Wave 1: includes original Wave 1 data from the time when Google, Facebook, and Grindr were advertising simultaneously and the data from the second phase of Wave 1.

N/A: not applicable.

## Secondary Outcomes

We explored the association of HIV test kit ordering and factors that could potentially affect ordering a test kit (see [Multimedia Appendix 3](#)). We found no statistically significant associations between test kit ordering and substance use, stage of health behavior change regarding HIV testing, and medical mistrust. However, ordering a HIV test kit was associated with the statement "People in my life would leave if I had HIV;" 48.1% (37/77) did not order a test kit whereas 33.7% (59/175) ordered a test kit;  $P=.04$ ). Participants who did not order a kit were more likely to agree with the statement "I think that new HIV/AIDS treatments can eradicate the virus from your body," compared to those who ordered a kit ( $P=.03$ ; Table (d) in [Multimedia Appendix 3](#)). People who ordered a self-test kit were more likely to disagree with the statement "I could not be friends with someone who has HIV/AIDS," compared to those who did not order a kit ( $P=.03$ ).

Of the 254 participants, 119 (46.8%) were classified as "high-risk alcohol use" and 67 (26.4%) as "problem alcohol use." Approximately 94 (37%) participants were classified as "high-risk cannabis use" and 19% of the participants as "problem cannabis use." Over half (136=53.5%) of the study participants reported that they were ready to start regularly testing for HIV ("Determination" stage of change), but only a small proportion of participants (12/177, 6.8%) among those who ordered a kit and 7.8% (6/77) of those who did not order a kit reported testing regularly ("Maintenance" stage of change). In total, 60 of the 254 (23.6%) participants agreed with the statement "I feel afraid of people living with HIV/AIDS" and only 9 (3.6%) agreed with the statement "I could not be friends with someone who has HIV/AIDS." Many participants believed

that mistakes are common in health care settings (155, 61.2%) and that organizations cover up their mistakes (153, 60.3%). They also reported being cautious toward health care organizations (151, 59.6%), with 159 (62.6%) feeling that patients have occasionally been misled or deceived by medical professionals.

Few participants (50/168, 29.8%) had a negative attitude toward taking PrEP ("I feel uncomfortable taking HIV medication when I don't have HIV."); some of them (21/165, 12.7%) were generally not ashamed to tell people ("I am ashamed to tell others that I am on PrEP."). However, they expressed concerns over the cost and long-term health effects. The reported barriers to starting PrEP included potential adverse effects of the medication (117/164, 71.3%) and fear of HIV treatment failure because of PrEP in case they get infected with HIV (138/163, 54.3%; Figure S2 in [Multimedia Appendix 3](#)). Facilitators to starting PrEP included the following: getting free HIV and sexually transmitted infection testing (134/162, 82.3%), acquiring free or low-cost PrEP (127/164, 77.4%), receiving a recommendation for PrEP from their doctor (119/163, 73%), and receiving additional counseling and support while on PrEP (118/163, 72.4%) (Figure S3 in [Multimedia Appendix 3](#)).

## Performance of Advertisement Campaigns

Throughout the duration of the promotional campaign, we spent approximately US \$20,000 in total per platform. Dating apps had the highest engagement (click-through rate of 4% resulting in 202 enrolled participants), even though they had the lowest number of impressions. Advertising through social media resulted in a high number of clicks (impressions) and low engagement (click-through rate of 0.6%). Information search sites recorded the lowest number of impressions among the 3 platform types and the lowest number of users who were

enrolled in the study (n=19), as shown in [Table 3](#). We calculated the cost per enrolled participant as US \$491.6 for social media, US \$88.8 for dating apps, and US \$841 for information search sites.

**Table 3.** Performance of advertisements by platform throughout the advertisement campaign in the National Institute on Drug Abuse Clinical Trials Network Social Media PrEP Study, 2020.

Platform	Impressions <sup>a</sup>	Clicks <sup>b</sup>	Click-through rate (%) <sup>c</sup>	Users screened <sup>d</sup>	Enrolled participants	Total funds spent (US \$)
Social media	3,864,778	21,399	0.6	2679	33	16,221.52
Dating apps	1,331,200	53,067	4	4390	202	17,939.40
Information search sites	708,770	10,869	1.5	2562	19	15,978.86

<sup>a</sup>Impressions refer to the number of times the advertisement is shown on a screen. A user may see the same advertisement multiple times.

<sup>b</sup>Clicks refer to number of times a user clicks on the advertisement.

<sup>c</sup>Click-through rate refers to the proportion of clicks or impressions.

<sup>d</sup>Users screened refers to users completing the screening survey.

## Discussion

### Principal Findings

In this study of MSM at risk for HIV infection, we investigated the effectiveness of promoting free home HIV self-test kits on various internet platforms. More than half of the participants ordered a self-test kit, although only a small proportion of HIV-negative individuals reported seeking PrEP services. Our results showed that dating apps were the most efficient platform to distribute HIV self-test kits to men at high risk for HIV infection. Risk behavior, attitudes toward HIV testing and treatment, perception of HIV-related stigma, and medical mistrust were not associated with ordering a self-test kit. Finally, we recorded high prevalence of alcohol and cannabis use among participants.

Overall, information search sites performed poorly in recruiting and enrolling individuals. The site advertisement metrics showed a better click-through rate than social media and a similar number of users screened, but ultimately only a small number of individuals enrolled in the study. Search engines have a broad audience as they are available to everyone with access to the internet, and they do not require an account. In comparison, dating apps had the highest click-through rate, screening numbers, and enrollment. Users of dating apps are more likely to be MSM and engage in high-risk behaviors, which could explain the higher engagement with the promoted study advertisements. Consequently, dating apps may be more cost-efficient in enrolling select individuals compared to other platforms. Using search engines for promotion may reach higher numbers of individuals, but dating apps achieved higher interaction with the promotion message in this study.

Another important difference between platforms that may have affected individual site performance is the type of advertisement message. Social media and dating apps use blast advertisements with images and text, whereas search engines use text-only promotional content. Researchers attempting to identify the best type of advertisement to reach MSM through the internet for free at-home HIV testing [29] showed that the click-through rate for a text-only advertisement on Google was 0.38%, whereas that for advertisements with images, such as the ones

used in social media and dating apps, was higher, between 0.77% and 2%.

There is a lack of published data regarding the performance of promotional campaigns to enroll participants or promote HIV prevention messages. This limits our capacity to make comparisons with similar campaigns. Our data showed that the cost of enrolling individuals from dating apps is lower compared to that for social media and information search sites. This is mainly due to the higher engagement and higher number of participants enrolled through dating apps. Future studies should collect and report advertisement campaign metrics as well as the costs of enrollment per participant screened and enrolled, which can allow for a better evaluation of the cost-effectiveness of different platforms.

### Secondary Findings

Our study demonstrated that HIV self-testing can reach individuals at high risk. We enrolled Latino and Black MSM at a high risk for HIV infection in 10 areas with a high incidence of HIV infection. The study population included individuals with inconsistent and infrequent condom use, and nearly 25% (64/254) of them reported that they had never tested for HIV. We also identified individuals who reported a preliminary positive result, which demonstrates the capacity of HIV home testing to reach hard-to-reach populations, overcome obstacles, and increase testing. Our findings underline the importance of identifying the best possible promotional platform that will allow public health programs to reach an even larger number of individuals at risk.

Our findings did not identify any major differences between participants who ordered a kit compared to those who did not order a test kit. However, our data showed a small statistical difference in terms of the questionnaires on self-perceived stigma, as well as the participant perceptions about the risks of HIV infection. Public health stakeholders should continue their efforts to educate individuals about HIV and support vulnerable individuals against stigma.

Substance use was common among study participants, especially alcohol and cannabis use. Similarly, Westmoreland et al [30] also reported a high incidence of cannabis use (55.8%) and alcohol use (22%) among a sample of MSM, transgender men,

and transgender women. Heavy alcohol use is associated with an increase in sexual behaviors that might put persons at risk for HIV acquisition and transmission [31]. Therefore, HIV prevention programs should include substance use screening and intervention services.

Medical mistrust has been associated with low intention of PrEP uptake [32,33] and poor medication adherence. Medical mistrust is also a barrier to HIV testing and causes disruptions in HIV care [34]. Study participants expressed a high level of mistrust toward medical providers and institutions. However, that did not seem to affect self-test kit ordering in our study. Additional research is needed to evaluate how medical mistrust may impact HIV testing and PrEP uptake.

Regarding PrEP, participants reported being informed of its benefits, comfortable taking PrEP, and not embarrassed about taking PrEP; however, they did report concerns about the adverse effects and the cost of PrEP. Similar concerns have been reported by Kota et al [35] in a cohort of MSM. Although PrEP is generally considered safe, public health messages should include more information about its low frequency of adverse effects and overall safety. Further awareness about access to low-cost PrEP might improve uptake and retention [34]. There are established state-sponsored programs that offer low-cost or free PrEP through in-person or telemedicine visits or with simple delivery via regular mail [36]. Additional efforts to promote those initiatives and programs in high-incidence areas, such as in the areas included in this study, may be necessary.

### Limitations

A few limitations should be taken into consideration when interpreting our findings. The study was conducted in 9 areas

with high HIV incidence; thus, the conclusions may not be generalizable to the whole country. Low enrollment and participation in waves affected our capacity to make broader comparisons between platforms and potentially between sites. In addition, we selected the most popular apps and sites as enrollment sites, grouping them into “platforms” with similar characteristics. Our goal was to investigate differences between the platforms. Thus, our findings are specific to the sites included in the campaigns.

### Conclusions

Our data suggest that certain dating apps may be an efficient way to reach young African American and Latinx MSM at high risk of HIV infection to promote the use of home HIV self-test kits. Dating apps are frequently used by many young MSM and offer a direct way of promoting HIV prevention to the target audiences. On the other hand, information search sites, such as Google, may require additional optimization for targeted messaging to be useful for HIV prevention. Results of this study could be used to inform public health agencies and stakeholders on what platforms are best to implement prevention campaigns.

New platforms, sites, and internet-based services are becoming available every day; therefore, research is necessary to evaluate the reach of public health and prevention campaigns using these new media outlets. Identifying and engaging individuals at increased risk for HIV infection in preventive care using entirely remote methods, including internet-based recruitment and remote access to preventive resources, is increasingly important and may represent the future of community-based HIV prevention.

---

### Acknowledgments

This manuscript reflects the views of the authors and may not reflect the opinions, views, and official policy or position of the US Department of Health and Human Services or any of its affiliated institutions or agencies. This work was supported by the National Institute on Drug Abuse (NIDA) Clinical Trials Network (grant UG1DA040309, Northeast Node), Data and Statistics Center (grants N271201400028C and 75N95019D00013), and Clinical Coordinating Center (grants N271201500065C and 75N95020D0001). NIDA had no role in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication. The content is solely the responsibility of the authors and does not necessarily represent the views of NIDA or the National Institutes of Health.

---

### Conflicts of Interest

SDY has received consulting funds from ElevateU, a startup company that has been funded by the National Institute on Drug Abuse to use digital outreach methods to recruit substance use participants. SDY is the principal investigator of gift funding from Facebook to the University of California. PJ and LMM were substantially involved in grant UG1DA040309, consistent with their role as scientific officers. The remaining authors have no conflicts of interest to declare.

---

#### Multimedia Appendix 1

Primary outcome analysis.

[[DOCX File, 24 KB - formative\\_v6i9e35648\\_app1.docx](#) ]

---

#### Multimedia Appendix 2

Sensitivity analyses.

[[DOCX File, 25 KB - formative\\_v6i9e35648\\_app2.docx](#) ]

## Multimedia Appendix 3

Secondary outcomes.

[\[DOCX File , 56 KB - formative\\_v6i9e35648\\_app3.docx \]](#)**References**

1. HIV surveillance report, 2018 (updated); vol.31. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/hiv/library/reports/hiv-surveillance/vol-31/index.html> [accessed 2022-08-20]
2. CDC HIV prevention progress report, 2019. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/hiv/pdf/policies/progressreports/cdc-hiv-preventionprogressreport.pdf> [accessed 2021-06-12]
3. MacGowan RJ, Chavez PR, Borkowf CB, Owen SM, Purcell DW, Mermin JH, eSTAMP Study Group. Effect of internet-distributed HIV self-tests on HIV diagnosis and behavioral outcomes in men who have sex with men: a randomized clinical trial. *JAMA Intern Med* 2020 Jan;180(1):117-125 [FREE Full text] [doi: [10.1001/jamainternmed.2019.5222](https://doi.org/10.1001/jamainternmed.2019.5222)] [Medline: [31738378](https://pubmed.ncbi.nlm.nih.gov/31738378/)]
4. Wray T, Chan PA, Simpanen E, Operario D. eTEST: developing a smart home HIV testing kit that enables active, real-time follow-up and referral after testing. *JMIR Mhealth Uhealth* 2017 May;5(5):e62 [FREE Full text] [doi: [10.2196/mhealth.6491](https://doi.org/10.2196/mhealth.6491)] [Medline: [28483744](https://pubmed.ncbi.nlm.nih.gov/28483744/)]
5. Huang E, Marlin RW, Young SD, Medline A, Klausner JD. Using Grindr, a smartphone social-networking application, to increase HIV self-testing among Black and Latino men who have sex with men in Los Angeles, 2014. *AIDS Educ Prev* 2016 Aug;28(4):341-350 [FREE Full text] [doi: [10.1521/aeap.2016.28.4.341](https://doi.org/10.1521/aeap.2016.28.4.341)] [Medline: [27427928](https://pubmed.ncbi.nlm.nih.gov/27427928/)]
6. Marlin RW, Young SD, Bristow CC, Wilson G, Rodriguez J, Ortiz J, et al. Piloting an HIV self-test kit voucher program to raise serostatus awareness of high-risk African Americans, Los Angeles. *BMC Public Health* 2014 Nov;14(1):1226 [FREE Full text] [doi: [10.1186/1471-2458-14-1226](https://doi.org/10.1186/1471-2458-14-1226)] [Medline: [25427749](https://pubmed.ncbi.nlm.nih.gov/25427749/)]
7. Hecht J, Sanchez T, Sullivan P, DiNunno E, Cramer N, Delaney K. Increasing access to HIV testing through direct-to-consumer HIV self-test distribution - United States, March 31, 2020-March 30, 2021. *MMWR Morb Mortal Wkly Rep* 2021 Sep;70(38):1322-1325 [FREE Full text] [doi: [10.15585/mmwr.mm7038a2](https://doi.org/10.15585/mmwr.mm7038a2)] [Medline: [34555001](https://pubmed.ncbi.nlm.nih.gov/34555001/)]
8. Self-testing: a strategy to improve access to HIV, viral hepatitis, and STI testing. NASTAD. 2021 Apr. URL: <https://nastad.org/resources/self-testing-strategy-improve-access-hiv-viral-hepatitis-and-sti-testing> [accessed 2022-06-19]
9. Patel RR, Few TE, Curoe K, Duff K, Burnside H, Miles G. A summary of HIV self-testing program models and lessons learned. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/hiv/pdf/testing/cdc-hiv-teesting-models-lessons-summary.pdf> [accessed 2022-08-20]
10. 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;5(3):e26 [FREE Full text] [doi: [10.2196/mhealth.7199](https://doi.org/10.2196/mhealth.7199)] [Medline: [28279949](https://pubmed.ncbi.nlm.nih.gov/28279949/)]
11. Fields EL, Hussen SA, Malebranche DJ. Mind the gap: HIV prevention among young Black men who have sex with men. *Curr HIV/AIDS Rep* 2020 Dec;17(6):632-642 [FREE Full text] [doi: [10.1007/s11904-020-00532-z](https://doi.org/10.1007/s11904-020-00532-z)] [Medline: [32914329](https://pubmed.ncbi.nlm.nih.gov/32914329/)]
12. Martinez O, Wu E, Shultz AZ, Capote J, López RJ, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. *J Med Internet Res* 2014 Apr;16(4):e113 [FREE Full text] [doi: [10.2196/jmir.3311](https://doi.org/10.2196/jmir.3311)] [Medline: [24763130](https://pubmed.ncbi.nlm.nih.gov/24763130/)]
13. Miller-Perusse M, Horvath KJ, Chavanduka T, Stephenson R. Recruitment and enrollment of a national sample of transgender youth via social media: experiences from Project Moxie. *Transgend Health* 2019 Jul;4(1):157-161 [FREE Full text] [doi: [10.1089/trgh.2018.0062](https://doi.org/10.1089/trgh.2018.0062)] [Medline: [31346546](https://pubmed.ncbi.nlm.nih.gov/31346546/)]
14. Rosengren AL, Huang E, Daniels J, Young SD, Marlin RW, Klausner JD. Feasibility of using Grindr to distribute HIV self-test kits to men who have sex with men in Los Angeles, California. *Sex Health* 2016 May;13(4):389-392 [FREE Full text] [doi: [10.1071/SH15236](https://doi.org/10.1071/SH15236)] [Medline: [27209076](https://pubmed.ncbi.nlm.nih.gov/27209076/)]
15. Guillory J, Wiant KF, Farrelly M, Fiacco L, Alam I, Hoffman L, et al. Recruiting hard-to-reach populations for survey research: using Facebook and Instagram advertisements and in-person intercept in LGBT bars and nightclubs to recruit LGBT young adults. *J Med Internet Res* 2018 Jun;20(6):e197 [FREE Full text] [doi: [10.2196/jmir.9461](https://doi.org/10.2196/jmir.9461)] [Medline: [29914861](https://pubmed.ncbi.nlm.nih.gov/29914861/)]
16. Iott BE, Veinot TC, Loveluck J, Kahle E, Golson L, Benton A. Comparative analysis of recruitment strategies in a study of men who have sex with men (MSM) in Metropolitan Detroit. *AIDS Behav* 2018 Jul;22(7):2296-2311. [doi: [10.1007/s10461-018-2071-z](https://doi.org/10.1007/s10461-018-2071-z)] [Medline: [29470808](https://pubmed.ncbi.nlm.nih.gov/29470808/)]
17. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and efficiency of online and offline recruitment methods: a web-based cohort study. *J Med Internet Res* 2017 Mar;19(3):e58 [FREE Full text] [doi: [10.2196/jmir.6716](https://doi.org/10.2196/jmir.6716)] [Medline: [28249833](https://pubmed.ncbi.nlm.nih.gov/28249833/)]
18. Edelstein Z, Salcuni P, Tsoi B, Kobrak P, Andaluz A, Katz D, et al. Feasibility and reach of a HIV self-test (HIVST) giveaway, New York City, 2015–2016. 2017 Presented at: Conference on Retroviruses and Opportunistic Infections (CROI); February 13-16, 2017; Seattle, WA p. 13-16 URL: <https://www.croiconference.org/abstract/feasibility-and-reach-home-test-giveaway-new-york-city-2015-2016/> [doi: [10.1097/olq.0000000000001144](https://doi.org/10.1097/olq.0000000000001144)]
19. Whiteman A, Wejnert C, Morris E, Burnett J. Using search engine data to explore interest in PrEP and HIV testing in the United States. *AIDS Behav* 2021 Mar;25(3):983-991. [doi: [10.1007/s10461-020-03057-z](https://doi.org/10.1007/s10461-020-03057-z)] [Medline: [33033997](https://pubmed.ncbi.nlm.nih.gov/33033997/)]

20. Johnson DC, Nobles AL, Caputi TL, Liu M, Leas EC, Strathdee SA, et al. Monitoring HIV testing and pre-exposure prophylaxis information seeking by combining digital and traditional data. *BMC Infect Dis* 2021 Feb;21(1):215 [FREE Full text] [doi: [10.1186/s12879-021-05907-0](https://doi.org/10.1186/s12879-021-05907-0)] [Medline: [33632140](https://pubmed.ncbi.nlm.nih.gov/33632140/)]
21. Parker JN, Hunter AS, Bauermeister JA, Bonar EE, Carrico A, Stephenson R. Comparing social media and in-person recruitment: lessons learned from recruiting substance-using, sexual and gender minority adolescents and young adults for a randomized control trial. *JMIR Public Health Surveill* 2021 Dec;7(12):e31657 [FREE Full text] [doi: [10.2196/31657](https://doi.org/10.2196/31657)] [Medline: [34855613](https://pubmed.ncbi.nlm.nih.gov/34855613/)]
22. Building healthy online communities recommendations for improving effectiveness of online ads to promote sexual health. BHOC. URL: [https://www.bhocpartners.org/wp-content/uploads/2019/10/BHOC-Purchaser-Meeting-Recommendations\\_10.22.19.pdf](https://www.bhocpartners.org/wp-content/uploads/2019/10/BHOC-Purchaser-Meeting-Recommendations_10.22.19.pdf) [accessed 2022-06-22]
23. Peters J. Grindr has been sold by its Chinese owner after the US expressed security concerns. *The Verge*. 2020 Mar. URL: <https://www.theverge.com/2020/3/6/21168079/grindr-sold-chinese-owner-us-cfius-security-concerns-kunlun-lgbtq> [accessed 2021-05-15]
24. Singer N, Krolik A. Grindr and OkCupid spread personal details, study says. *New York Times*. 2020 Jan. URL: <https://www.nytimes.com/2020/01/13/technology/grindr-apps-dating-data-tracking.html> [accessed 2021-05-15]
25. Lemley SM, Klausner JD, Young SD, Stafylis C, Mulatya C, Oden N, et al. Comparing web-based platforms for promoting HIV self-testing and pre-exposure prophylaxis uptake in high-risk men who have sex with men: protocol for a longitudinal cohort study. *JMIR Res Protoc* 2020 Oct;9(10):e20417 [FREE Full text] [doi: [10.2196/20417](https://doi.org/10.2196/20417)] [Medline: [33074164](https://pubmed.ncbi.nlm.nih.gov/33074164/)]
26. Hartz SM. Capsule commentary on Gryczynski et al., Validation of the TAPS-1: a four-item screening tool to identify unhealthy substance use in primary care. *J Gen Intern Med* 2017 Sep;32(9):1026 [FREE Full text] [doi: [10.1007/s11606-017-4119-6](https://doi.org/10.1007/s11606-017-4119-6)] [Medline: [28721534](https://pubmed.ncbi.nlm.nih.gov/28721534/)]
27. Stolte IG, Dukers NH, Geskus RB, Coutinho RA, de Wit JBF. Homosexual men change to risky sex when perceiving less threat of HIV/AIDS since availability of highly active antiretroviral therapy: a longitudinal study. *AIDS* 2004 Jan;18(2):303-309. [doi: [10.1097/00002030-200401230-00021](https://doi.org/10.1097/00002030-200401230-00021)] [Medline: [15075549](https://pubmed.ncbi.nlm.nih.gov/15075549/)]
28. Chiu YF. Multiple comparisons and multiple tests. Using the SAS system. Peter H. Westfall, Randall D. Tobias, Dror Rom, Russell D. Wolfinger and Yosef Hochberg, SAS Institute, Cary, U.S.A. 2000. No. of pages: xiv + 397. Price: DKK 412.00. ISBN 1-58025-397-0. *Statist Med* 2002 May;21(10):1499-1500. [doi: [10.1002/sim.1168](https://doi.org/10.1002/sim.1168)]
29. Grov C, Stief M, Westmoreland DA, MacCrate C, Mirzayi C, Nash D, Together 5000 Study Team. Maximizing response rates to ads for free at-home HIV testing on a men-for-men geosocial sexual networking app: lessons learned and implications for researchers and providers. *Health Educ Behav* 2020 Feb;47(1):5-13 [FREE Full text] [doi: [10.1177/1090198119893692](https://doi.org/10.1177/1090198119893692)] [Medline: [31896287](https://pubmed.ncbi.nlm.nih.gov/31896287/)]
30. Westmoreland DA, Carrico AW, Goodwin RD, Pantalone DW, Nash D, Grov C. Higher and higher? Drug and alcohol use and misuse among HIV-vulnerable men, trans men, and trans women who have sex with men in the United States. *Subst Use Misuse* 2021 Nov;56(1):111-122 [FREE Full text] [doi: [10.1080/10826084.2020.1843057](https://doi.org/10.1080/10826084.2020.1843057)] [Medline: [33153358](https://pubmed.ncbi.nlm.nih.gov/33153358/)]
31. Woolf SE, Maisto SA. Alcohol use and risk of HIV infection among men who have sex with men. *AIDS Behav* 2009 Aug;13(4):757-782. [doi: [10.1007/s10461-007-9354-0](https://doi.org/10.1007/s10461-007-9354-0)] [Medline: [18236149](https://pubmed.ncbi.nlm.nih.gov/18236149/)]
32. Brooks RA, Allen VC, Regan R, Mutchler MG, Cervantes-Tadeo R, Lee S. HIV/AIDS conspiracy beliefs and intention to adopt preexposure prophylaxis among black men who have sex with men in Los Angeles. *Int J STD AIDS* 2018 Mar;29(4):375-381 [FREE Full text] [doi: [10.1177/0956462417727691](https://doi.org/10.1177/0956462417727691)] [Medline: [28853676](https://pubmed.ncbi.nlm.nih.gov/28853676/)]
33. Kimball D, Rivera D, Gonzales M, Blashill AJ. Medical mistrust and the PrEP cascade among Latino sexual minority men. *AIDS Behav* 2020 Dec;24(12):3456-3461 [FREE Full text] [doi: [10.1007/s10461-020-02916-z](https://doi.org/10.1007/s10461-020-02916-z)] [Medline: [32405726](https://pubmed.ncbi.nlm.nih.gov/32405726/)]
34. Johnson LM, Green HD, Koch B, Harding R, Stockman JK, Wagner KD. Correlates of medical mistrust among minority women at risk for HIV and their networks. *Health Educ Behav* 2021 Dec;48(6):860-872 [FREE Full text] [doi: [10.1177/1090198120986783](https://doi.org/10.1177/1090198120986783)] [Medline: [33622086](https://pubmed.ncbi.nlm.nih.gov/33622086/)]
35. Kota KK, Mansergh G, Stephenson R, Hirshfield S, Sullivan P. Sociodemographic characteristics of HIV pre-exposure prophylaxis use and reasons for nonuse among gay, bisexual, and other men who have sex with men from three US cities. *AIDS Patient Care STDS* 2021 May;35(5):158-166 [FREE Full text] [doi: [10.1089/apc.2020.0261](https://doi.org/10.1089/apc.2020.0261)] [Medline: [33901403](https://pubmed.ncbi.nlm.nih.gov/33901403/)]
36. Ko JS, Stafylis C, Klausner JD. Mobile health promotion of human immunodeficiency virus self-testing in the United States. *Mhealth* 2020 Jan;6:10 [FREE Full text] [doi: [10.21037/mhealth.2019.10.05](https://doi.org/10.21037/mhealth.2019.10.05)] [Medline: [32190621](https://pubmed.ncbi.nlm.nih.gov/32190621/)]

---

## Abbreviations

- MSM:** men who have sex with men  
**PrEP:** pre-exposure prophylaxis
-

*Edited by A Mavragani; submitted 12.12.21; peer-reviewed by R McGowan, A Algarin; comments to author 10.05.22; revised version received 11.07.22; accepted 10.08.22; published 23.09.22.*

*Please cite as:*

*Stafylis C, Vavala G, Wang Q, McLeman B, Lemley SM, Young SD, Xie H, Matthews AG, Oden N, Revoredo L, Shmueli-Blumberg D, Hichborn EG, McKelle E, Moran LM, Jacobs P, Marsch LA, Klausner JD*

*Relative Effectiveness of Social Media, Dating Apps, and Information Search Sites in Promoting HIV Self-testing: Observational Cohort Study*

*JMIR Form Res 2022;6(9):e35648*

*URL: <https://formative.jmir.org/2022/9/e35648>*

*doi: [10.2196/35648](https://doi.org/10.2196/35648)*

*PMID: [36149729](https://pubmed.ncbi.nlm.nih.gov/36149729/)*

©Chrysovalantis Stafylis, Gabriella Vavala, Qiao Wang, Bethany McLeman, Shea M Lemley, Sean D Young, Haiyi Xie, Abigail G Matthews, Neal Oden, Leslie Revoredo, Dikla Shmueli-Blumberg, Emily G Hichborn, Erin McKelle, Landhing M Moran, Petra Jacobs, Lisa A Marsch, Jeffrey D Klausner. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 23.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Comparing Professional and Consumer Ratings of Mental Health Apps: Mixed Methods Study

Georgie Hudson<sup>1,2,3</sup>, BSc; Esther Negbenose<sup>1,2</sup>, BSc; Martha Neary<sup>4</sup>, BA, MSc; Sonja M Jansli<sup>1,2</sup>, BSc, MSc; Stephen M Schueller<sup>4,5</sup>, BA, MA, PhD; Til Wykes<sup>1,2</sup>, PhD; Sagar Jilka<sup>1,2,6</sup>, BSc, MRes, PhD

<sup>1</sup>Institute of Psychiatry Psychology and Neuroscience, King's College London, London, United Kingdom

<sup>2</sup>South London and Maudsley NHS Foundation Trust, London, United Kingdom

<sup>3</sup>Division of Psychology and Language Sciences, University College London, London, United Kingdom

<sup>4</sup>Department of Psychological Science, University of California, Irvine, CA, United States

<sup>5</sup>Department of Informatics, University of California, Irvine, CA, United States

<sup>6</sup>Warwick Medical School, University of Warwick, Coventry, United Kingdom

**Corresponding Author:**

Sagar Jilka, BSc, MRes, PhD

Warwick Medical School

University of Warwick

Medical School Building

Coventry, CV4 7AL

United Kingdom

Phone: 44 7708715627

Email: [sagar.jilka@warwick.ac.uk](mailto:sagar.jilka@warwick.ac.uk)

## Abstract

**Background:** As the number of mental health apps has grown, increasing efforts have been focused on establishing quality tailored reviews. These reviews prioritize clinician and academic views rather than the views of those who use them, particularly those with lived experiences of mental health problems. Given that the COVID-19 pandemic has increased reliance on web-based and mobile mental health support, understanding the views of those with mental health conditions is of increasing importance.

**Objective:** This study aimed to understand the opinions of people with mental health problems on mental health apps and how they differ from established ratings by professionals.

**Methods:** A mixed methods study was conducted using a web-based survey administered between December 2020 and April 2021, assessing 11 mental health apps. We recruited individuals who had experienced mental health problems to download and use 3 apps for 3 days and complete a survey. The survey consisted of the One Mind PsyberGuide Consumer Review Questionnaire and 2 items from the Mobile App Rating Scale (star and recommendation ratings from 1 to 5). The consumer review questionnaire contained a series of open-ended questions, which were thematically analyzed and using a predefined protocol, converted into binary (positive or negative) ratings, and compared with app ratings by professionals and star ratings from app stores.

**Results:** We found low agreement between the participants' and professionals' ratings. More than half of the app ratings showed disagreement between participants and professionals (198/372, 53.2%). Compared with participants, professionals gave the apps higher star ratings (3.58 vs 4.56) and were more likely to recommend the apps to others (3.44 vs 4.39). Participants' star ratings were weakly positively correlated with app store ratings ( $r=0.32$ ,  $P=.01$ ). Thematic analysis found 11 themes, including issues of user experience, ease of use and interactivity, privacy concerns, customization, and integration with daily life. Participants particularly valued certain aspects of mental health apps, which appear to be overlooked by professional reviewers. These included functions such as the ability to track and measure mental health and providing general mental health education. The cost of apps was among the most important factors for participants. Although this is already considered by professionals, this information is not always easily accessible.

**Conclusions:** As reviews on app stores and by professionals differ from those by people with lived experiences of mental health problems, these alone are not sufficient to provide people with mental health problems with the information they desire when choosing a mental health app. App rating measures must include the perspectives of mental health service users to ensure ratings represent their priorities. Additional work should be done to incorporate the features most important to mental health service users into mental health apps.

## KEYWORDS

well-being; apps; patient and public involvement; coproduction; mental health; service user; technology; mobile health; mHealth; digital; mobile phone

## Introduction

### Interest in Mental Health Apps

Digital technologies can expand access to mental health care. The availability of smartphone apps to support mental health and well-being has increased over the last few years, with some evidence supporting their use in depression [1], anxiety [2], and other mental health conditions [3]. A survey of interest in smartphone apps among military veterans found that 43% indicated an interest in using a mental health app [4]; however, despite this enthusiasm, only 11% had done so, with a major barrier to adoption including concerns around a lack of proof of efficacy (72%). This lack of proof is reflected in the rapid expansion of available mental health apps with little regulation or oversight [5-9]. This problem of “high availability but low evidence base” [10] means that many publicly available products have little or no evidence supporting their use.

### Consumer Preferences in Mental Health Apps

Consumers want access to clear information when choosing an app [11,12], and several measures have been developed to provide this information on the app of choice. The most commonly used measure is the Mobile App Rating Scale (MARS) [13], which assesses *engagement*, *aesthetics*, and *usability*. However, on its own, this measure does not provide enough information to allow service users to decide whether to use an app [14]. ORCHA, a for-profit company [15], reviews apps on *current standards*, *regulations*, and *good practices* but only provides a composite score, which does not allow service users to identify factor ratings that are most personally important to their choice. When consumers are choosing apps, often the only measures available are star ratings and reviews. These reviews may be written by genuine app users, but there are large numbers of fake reviews, which are hard to distinguish from genuine ones [16]. In the United Kingdom, the general population had increased experiences of insomnia, anxiety, low mood, and general psychological distress during the COVID-19 pandemic and subsequent lockdowns [17-19]. There was also a 200% increase in mental health app use [20], with many people relying on mental health care apps on their phones when their usual care was disrupted. Owing to this increased use, we need to understand what matters to consumers when selecting mental health apps to convey that information so they can make informed choices. A study found that people with mental health needs value aesthetics and data security when choosing a mental health app [21]; however, it is not known whether these values differ from professionals' views.

### This Study

This study aimed to fill this gap by understanding how those with lived experience perceive mental health apps and how these views differ from clinicians' and academics' reviews and those provided on app stores. The findings will contribute to a

broader understanding of consumers' use of and opinions on mental health apps.

## Methods

### Design

This was a cross-sectional mixed methods study evaluating 11 mental health apps.

### Ethics Approval

The study received ethical approval from the King's College London Psychiatry, Nursing and Midwifery Research Ethics Subcommittees on November 30, 2020 (LRS-20/21-21137). All participants provided written consent before participating.

### Patient and Public Involvement

We consulted the Young Person's Mental Health Advisory Group [22] on the design of the study. This included the length of time participants should use each app and phrasing of questionnaire items to improve clarity.

### Participants and Recruitment

Participants were recruited using volunteer sampling from advisory groups, local mental health groups, and a general university-wide newsletter. Participants were included if they were aged  $\geq 18$  years, living in the United Kingdom, had access to a smartphone, were able to download smartphone apps, and had a history of mental health problems. Participants were not screened for psychiatric diagnoses but were asked whether they had “experience of mental health difficulties.” This was judged to be most appropriate in this study, as smartphone use is lower in those with serious mental illnesses [23,24], and most mental health apps target improving general well-being rather than severe symptoms.

### Apps

#### App Selection

We initially selected 12 apps for the trial. All apps had to be freely available for download, as consumers strongly prefer free apps [21]. Overall, 50% (6/12) of our initial apps were selected from the highest-ranking mental health apps on Google Play and iOS app stores, and these were supplemented with 6 of the highest-rated apps based on the One Mind PsyberGuide Credibility Rating Scale [25]. We removed 1 app as it required users to sign up for a free trial that would automatically upgrade to an annual subscription, reducing the number of apps to 11.

#### Overview of Apps

The 11 apps assessed were Breethe, Calm, Headspace, Insight Timer Meditation, MindDoc, MindShift, Reflectly, Remente, Sanvello, Self-Help for Anxiety, and Woebot. These apps can be divided into three categories: meditation (Calm, Headspace, Breethe, and Insight Timer Meditation), journaling (Reflectly,

Remente, and MindDoc), and cognitive behavioral therapy (Woebot, MindShift, Sanvello, and Self-Help for Anxiety). Despite these categories, apps have a considerable overlap of functionality, as demonstrated in the study by Lagan et al [26]. For example, although based on cognitive behavioral therapy, MindShift, Sanvello, and Self-Help for Anxiety also use meditation and relaxation techniques. Therefore it is not possible to differentiate reviews based on app category.

## Professional Reviewers

Professional reviews were completed by a team of 4 highly trained raters. Training takes between 3 and 4 hours and consists of a video tutorial and app ratings, followed by in-person reliability checks against an expert rater.

## Measures

**Textbox 1** presents the participant and professionals' measures and app store ratings.

**Textbox 1.** Participants' and professionals' measures and app store ratings.

### Participants' measures

- *Demographic information and app use:* Age, ethnicity, gender, education, employment status, average daily digital device use, and most frequently used digital device.
- *One Mind PsyberGuide Consumer Review Questionnaire:* 12 open-ended questions were derived from the metrics used by One Mind PsyberGuide, including the One Mind PsyberGuide credibility scale [25], Mobile App Rating Scale (MARS) [13], and One Mind PsyberGuide transparency scale [27]. These 12 questions were mapped onto six app domains:
  1. Ease of use ("How easy or hard was this app to use?")
  2. Difficulties of use ("Were there any parts of the app that were confusing or difficult to use?")
  3. Engagement ("Did you enjoy using this app?")
  4. Aesthetics ("What did you think about how this app looked?")
  5. Perceived impact on well-being ("What impact, if any, did this app have on your well-being?")
  6. Data security ("Did you feel confident that the data you entered in this app was secure?")
- *MARS [13]:* We used two items from the MARS based on recommendations by our service user advisers:
  1. *Recommendation ratings:* If they would recommend this app to people who might benefit from it on a 5-point Likert scale.
  2. *MARS star ratings:* A star rating from 1 to 5.

### Professionals' measures

Professionals' ratings of all apps were collected from the One Mind PsyberGuide website. The data were as follows:

- *MARS [13]:* Professionals' recommendations and MARS star ratings from 1 to 5 were assessed. This measure also captured professionals' ratings of the domains of app functionality, engagement, aesthetics, and perceived impact on well-being, which were mapped onto the participants' ratings. These were measured using a 5-point Likert scale.
- *PsyberGuide Transparency Score:* Professionals' ratings of the presence and quality of a privacy policy were used. This measure comprises 7 subquestions and results in a binary classification of data security (acceptable or unacceptable). This measure was used previously [27] and adapted from the Enlight evaluation tool [28].

### App store ratings

- Average star ratings for each app were collected from both iOS and Google Play stores on November 19, 2021, and the scores were averaged across both app stores.

## Procedure

Once participants had consented, they were randomly allocated 3 of the 11 apps. They used the 3 apps over 3 days, with a total participation period of 3 days, as suggested by our service user advisers. This also corroborates previous work, which found that the number of times a mental health app is opened declines by 80% over the first 10 days of use [29]. Participants were encouraged to explore the features of the apps and to use the apps for 10 to 60 minutes per day, spending an equal amount of time on each app. On the evening of the third day, participants completed the MARS ratings and the One Mind PsyberGuide Consumer Review for each app via SurveyMonkey.

We compared reviews of consumers and professionals on the following six domains:

1. Ease of use ("How easy or hard was this app to use?")
2. Difficulties of use ("Were there any parts of the app that were confusing or difficult to use?")
3. Aesthetics ("What do you think about how this app looked?")
4. Engagement ("Did you enjoy using the app? (eg, was it engaging, fun, or boring?)")
5. Perceived impact on well-being ("What impact, if any, did this app have on your well-being?")

6. Data security (“Did you feel confident that the data you entered in this app were secure? Why, or why not?”)

## Data Analysis

### Quantitative

We converted the qualitative text from the One Mind PsyberGuide Consumer Review into a quantitative binary classification (“1” a positive experience and “0” a negative experience) using a predefined protocol ([Multimedia Appendix](#)

1). Two researchers independently conducted this coding (GH and SMJ), and any disagreements were resolved via discussion with 2 other independent researchers (SJ and TW) to provide a final quantitative classification for each rating across all apps and domains.

Median scores for professional reviews were calculated for each of the 6 domains. Any scores on or below the median were negative (score 0). [Table 1](#) presents the participant and professional scores.

**Table 1.** Participant and professional scores and their interpretations.

	Participant=0	Participant=1
Professional=0	Negative agreement (both participants and professionals rate negatively)	Professional negative (participants rate positively but professionals rate negatively)
Professional=1	Participant negative (participants rate negatively but professionals rate positively)	Positive agreement (both participants and professionals rate positively)

The primary outcome measure was “participant negative” (participant rated negatively but professional rated positively) across the 6 domains. In addition, we tested participant-professional agreement using the weighted Cohen  $\kappa$  statistic for recommendation ratings and MARS star ratings for all apps for which these ratings were available. Furthermore, 2 PsyberGuide professionals’ recommendations and MARS star ratings were available for each app; therefore, we report the comparison of participant ratings against each professional and an average. We report the Cohen [30] interpretations of the  $\kappa$  values (0.01-0.2, none to slight; 0.21-0.4, fair; 0.41-0.6, moderate; 0.61-0.8, substantial; and 0.81-1, almost perfect agreement). As we asked the participants 2 questions relating to the functionality of the apps (ease of use and difficulties of use), these ratings were both compared with the professionals’ functionality score on PsyberGuide. Professional recommendations and MARS star ratings were not available for Self-Help for Anxiety.

MARS star ratings were compared with app store ratings from the iOS app store and Google Play using Pearson correlations to compare genuine users with lived experiences with app store reviews. All quantitative analyses were performed using SPSS version 27 (IBM Corp) for Windows [31].

### Qualitative

All open-ended survey responses were thematically analyzed using the Braun and Clarke [32] method, which was also used

in previous publications [33,34]. Themes were inductively extracted by 2 researchers (GH and EN) independently, using the analysis framework by Pope et al [35]. This is a 5-stage process and involves (1) familiarizing with raw data, (2) identifying a thematic framework, (3) indexing, (4) charting, and (5) mapping and interpreting—defining concepts, mapping the range and nature of phenomena, and creating typologies. Each of the 2 researchers independently and inductively coded all the participant responses, resulting in 2 thematic frameworks. The 2 researchers then created the final inductive framework together by discussing the similarities and differences between the 2 frameworks and using the elements of the multiple coding approach [36]. Theme names were decided collaboratively by the 2 researchers. Any discrepancies were resolved through discussion between the 2 researchers and were independently checked by a third researcher (SMJ). All qualitative analyses were performed using NVivo version 12 for Windows [37].

## Results

### Sample Characteristics

A total of 21 people participated in the study. Most were women and educated to a degree level, but they were ethnically diverse. [Table 2](#) presents the breakdown of participant characteristics.

**Table 2.** Sample characteristics (N=21).

Characteristic	Participants
<b>Gender, n (%)</b>	
Female	15 (71)
Male	5 (24)
Nonbinary	1 (5)
Age (years), mean (SD, range)	29.10 (11.01, 20-60)
<b>Ethnicity, n (%)</b>	
Asian or Asian British	7 (33)
Black or Black British	4 (19)
White British	6 (29)
Other	4 (24)
<b>Education status, n (%)</b>	
No qualifications	1 (45)
A-level	7 (33)
Degree	12 (57)
PhD	1 (5)
<b>Employment status, n (%)</b>	
Employed (full or part-time)	10 (48)
Student	6 (29)
Unemployed	3 (14)
Retired	1 (5)
Receiving ESA <sup>a</sup>	1 (5)
Previous use of well-being apps, n (%)	13 (62)
<b>Average daily digital device use, n (%)</b>	
<1 hour	2 (10)
1-3 hours	3 (14)
3-5 hours	7 (33)
>5 hours	9 (43)
<b>Most frequently used digital device type, n (%)</b>	
Smartphone	13 (62)
Desktop	8 (38)

<sup>a</sup>ESA: employment and support allowance.

### Do Participants Agree With Professionals' Reviews?

Overall, there was little agreement between the participants' and professionals' reviews (Table 3), with most app ratings classified as disagreements (53.2% vs 46.8% agreements).

Participants were much less positive about the apps than professionals, with difficulties in use being the most different (Table 4). Multimedia Appendix 2 gives a more detailed account.

**Table 3.** Overall number and relative percentage of negative agreement, positive agreement, professional negative, and participant negative (N=372).

Particulars	Participant=0, n (%)	Participant=1, n (%)
Professional=0	Negative agreement <sup>a</sup> , 76 (20.4)	Professional negative <sup>b</sup> , 149 (40.1)
Professional=1	Participant negative <sup>c</sup> , 49 (13.2)	Positive agreement <sup>d</sup> , 98 (26.3)

<sup>a</sup>Both participants and professionals rate negatively.

<sup>b</sup>Participants rate positively but professionals rate negatively.

<sup>c</sup>Participants rate negatively but professionals rate positively.

<sup>d</sup>Both participants and professionals rate positively.

**Table 4.** The breakdown (number and percentage) of which domains participants scored negatively and professionals scored positively (“participant negatives”; N=49).

Domain	Participant negatives, n (%)
Difficulties of use	17 (27)
Data security	10 (16)
Aesthetics	9 (15)
Ease of use	7 (11)
Perceived impact on well-being	4 (67)
Engagement	2 (3)

### Do Participants Agree With Professionals’ Views?

There was *moderate to substantial* agreement between the 2 professionals’ recommendation ratings (Cohen  $\kappa_w=0.667$ ;  $P=.008$ ) and MARS star ratings (Cohen  $\kappa_w=0.571$ ;  $P=.008$ ). However, there was little (*none to slight*) agreement between the participants’ and professionals’ recommendation ratings (Cohen  $\kappa_w=0.048$ ; professional 1, Cohen  $\kappa_w=0.047$ ; professional 2, Cohen  $\kappa_w=0.048$ ). Participants gave lower recommendation ratings on average (mean 3.44, SD 1.09) than the 2 professionals (professional 1: mean 4.22, SD 1.30; professional 2: mean 4.56, SD 1.01). There was also little (*none to slight*) agreement between participants’ and professionals’ MARS star ratings (Cohen  $\kappa_w=0.108$ ; professional 1, Cohen  $\kappa_w=0.124$ ; professional 2, Cohen  $\kappa_w=0.092$ ), with participants again giving lower star ratings on average (mean 3.58, SD 0.91) than the 2 professionals (professional 1: mean 4.44, SD 0.73; professional 2: mean 4.67, SD 0.71).

### Do Participants Agree With App Store Ratings?

Participants’ MARS star ratings of apps were significantly positively correlated with average app store ratings ( $r=0.32$ ;  $P=.01$ ) and with individual iOS app store ( $r=0.27$ ;  $P=.04$ ) and Google Play ( $r=0.31$ ;  $P=.02$ ) ratings. These correlations were low, despite the agreement between iOS app store and Google Play ratings ( $r=0.73$ ;  $P<.001$ ).

### What Do Participants Want From Mental Health Apps?

The thematic analysis of participants’ qualitative data found 11 themes.

#### Cost

One of the most common themes mentioned by respondents was cost. This was largely in response to the survey item “what

did you like the least about this app?” All users were able to engage in some content without paying but found it “frustrating to see so many options which you can’t use due to having the free version,” especially when “it wasn’t allowing me to experiment with things and find what’s right for me before purchasing.” Therefore, the most frustrating part of the experience was the hidden costs introduced by freemium or other forms of pricing that provide a limited experience of the app for free, with features behind a paywall, which did not allow participants to try these features without paying. Many participants reported that on the free versions of the apps, there were many adverts, often for users to “upgrade to premium,” which participants found “excessive” and “would ruin the flow or calm I had going.”

#### Aesthetics

The user interface contributed to people’s enjoyment. The structure of app features and layout for each section were the main influencing factors. For example, Insight Timer was described as having a “*professional*” layout, which made the app appear more legitimate. Another was described as clumsy and inconsistent in design, which made it look “like it’s in beta format.” Some colors made the app more engaging, with others providing a “nice and calming” feeling. Some users also felt that their “screen appeared quite crowded due to the number of options,” reducing the appeal.

#### Ease of Use, Navigation, and Functionality

Most apps were described as “very easy and simple” to use and they were able to navigate features “without having to try very hard to find them.” However, prompts to guide navigation to specific features are necessary, especially for users with no prior experience of using well-being apps. Features on the app should also load quickly so users are not irritated. Some apps had technological problems such as glitches, where “one time where

it lost my entire journal entry.” This could make the apps difficult for participants to use as “sometimes there [w]as content there and at other times, it said ‘We couldn’t find any results.’ I found this quite frustrating.”

### ***Interactivity***

Participants valued interactivity and particularly enjoyed “daily quotes” and “a voice assistant.” However, interactive prompts, reminders, and notifications also garnered mixed opinions. While some believed the prompts were “useful...for people on their anxiety recovery journey” and “made me actually check in with how I was doing,” others found them unnecessary and “had to turn them off because they got annoying.” These participants wanted the app notification frequency to be optimized so they do not feel they are being nagged.

### ***Personalization and Customization***

There were mixed responses on the extent to which apps allowed users to have a customized experience. Most participants believed it was an important feature of an enjoyable app experience. Some apps used natural language processing to provide relevant and appropriate responses to user input. Participants, therefore, received “personal insights tailored to you[r] mental health.” The option of customizing color schemes and audio voices was highly favored. Personalization also allowed users to “save your favourite content and create playlists,” “so can reuse particularly helpful courses.”

### ***Education and Teaching***

Many apps included information in the form of articles and blog posts. Users repeatedly mentioned that this content provided “useful education on mental health issues,” specifically those related to anxiety and depression. Information on managing symptoms and coping mechanisms was especially useful. Participants reported that this educational content helped them understand their own mental health, “[teaching] me more about background of anxiety” and proving participants with “more of a technical understanding of anxiety and stress.” However, it was suggested that this information was not always at a sufficiently deep level of complexity for those with a diagnosis of depression or anxiety (“for someone who already knows about this then it would probably be quite basic”).

### ***Tracking and Goal Setting***

Apps that provide a clear way to set goals and track progress were said to be interesting and useful. Tracking progress helped participants to “track your mood and identify possible triggers,” and monitoring their mental health “helps me monitor my mind and what triggers me.” Participants found the ability to track anxiety symptoms specifically helpful as it “allows you to understand your anxiety and how it progresses on the graph so you can track progress on the anxiety tracker.” However, other users reported that tracking features caused negative emotions because it was not “helpful or productive to...see the amount of days I potentially felt low.”

### ***Variety of Features***

Users appreciated a combination of audio and visual content. Apps that require users to perform daily tasks should ensure that those tasks are not repetitive, and some apps should “add

more features to attract users.” This often included further developing the existing features (“the check in feature didn’t have enough guidance and was quite bare bones and didn’t provide counter to negative thoughts for instance”). Two participants also requested more features “aimed at the teenage community,” “as there is enough material for adults and kids but I didn’t see as much as I was hoping for teens.” However, a variety of features can be excessive— “sometimes difficult to choose which activity to focus on because there was too much content.” “The huge amount of content stopped it being engaging,” and it was recommended that “they should focus on a few core features.”

### ***Data Security and Privacy Concerns***

The general sentiment among the respondents was that they had no concerns about the security of their data. This was mainly because they did not enter “anything overly personal.” Others were able to register via their Apple or Google Play account which was perceived as a legitimate process of verification and data protection. Some relied on their prior understanding of UK General Data Protection Regulation regulations to determine the security of their data (“they would have to be complying with the law”). Some reported that although they had seen the declaration on the apps, they thought it should be better signposted (“perhaps a disclaimer can be added to the start of the app making it clear about data security”).

### ***Integration With Daily Life***

Well-being apps can be used to support users at various points of the day and various locations. Many users were able to schedule times to use them at their own convenience (“specific meditations designed for different times of day—starting the day, commuting, focusing at work”). They also appreciated content relevant to their specific life circumstances (“had a section on dealing with corona which was very useful” and “some exercises to help you cope with specific aspects of life”). Although some participants found “the exercises were nice and short which is very convenient for someone with an insanely busy life like mine,” others found that “as it’s very time consuming, regular use of the app may not [be] sustainable for the long term.”

### ***Impact on Well-being***

Users reported a change in their well-being, specifically helping people “feel significantly less anxious.” Others appreciated the exercises and courses they engaged with as they were thought-provoking and promoted introspection. Information specifically about mental health status and how to use different strategies to cope with adverse experiences was welcomed. The guided journeys aided them in reducing maladaptive thought processes by helping them understand the origin of their negative thinking patterns which, in turn, helped reduce feelings of anxiety and low mood. Respondents were provided with “exercises to cope with these feelings but also knowledge to understand what anxiety is,” which were greatly appreciated.

## Discussion

### Principal Findings

To the best of our knowledge, this is the first study to investigate how reviews of mental health apps by professionals differ from those by people with mental health problems. Most reviews focus exclusively on professionals' opinions [38-40], and reviews that appear to be from genuine app users, such as on app stores, are often false [16]. We have demonstrated that these opinions differ, and therefore, professional reviews and those on app stores are not sufficient to provide those with mental health problems the information they want when selecting a mental health app.

We found low levels of agreement among the ratings of professionals, app stores, and people with mental health problems. Participants placed a great deal of importance on app functionality, and most themes generated through the qualitative analysis were related to this aspect. They appreciated a variety of features, which were easy to use, interactive, and with the capacity for personalization [41]. Aesthetics were also very important, as our participants emphasized the importance of a professional layout, with engaging colors and a simple structure. The highest number of participant negatives was for the domain "difficulties of use," suggesting that current professional ratings are overestimating the ease with which the apps can be used. Overall, we found that more than half (53%) of the app ratings showed disagreements between participants and professionals. This high level of disagreement shows that professionals have highly different views of what is important in a mental health app, compared with those with personal experience of mental health problems.

We found weak positive correlations between app store and participant ratings. This low agreement suggests that ratings of app stores are not representative of the opinions of those who have experienced mental health problems, and therefore, app stores are not sufficient to provide the information desired by those with mental health problems. This discrepancy may be due to the high number of fake reviews in app stores [16], or it may reflect differing priorities between those with lived experiences of mental health problems and laypeople in the general population. We suggest that it would be beneficial for people with lived experiences to rate mental health apps, rather than exclusively professionals, to ensure the ratings are more accurate and representative of mental health service users' opinions.

### Comparisons With Prior Work

The variety of features our participants preferred mirrors other studies, such as a scoping review of 37 studies on mental health chatbots, which found that usefulness and ease of use were the most frequently assessed features [35]. Importantly, we found that professionals and those with mental health problems disagree in their ratings of mental health apps. This aligns with previous research findings that participants could independently complete less than half of the tasks in apps targeted at chronic conditions [42] and expressed significant frustration with the design features and navigation of the apps. By engaging with those with lived experiences, app designers and professional

raters can identify the features of apps that are most important to this population.

Professional raters may also miss some domains that users with lived experiences emphasized. The ability to track and measure their mental health, as well as the provision of informative articles about mental health, was praised by the participants. This replicates other studies. For example, almost three-quarters of people (from a sample where half had experienced mental illness) perceived monitoring or showing progress toward a goal as useful in a mental health app [43]. Our participants' dislike of excessive notifications was also mirrored in the study by Thornton and Kay-Lambkin [43]. Cost was one of the most frequently mentioned negative aspects of apps, highlighting a major issue with accessibility and inclusivity. Professional reviews on PsyberGuide frequently consider cost in their narrative reviews; however, it is not incorporated in their numerical ratings. Thus, consumers may be influenced by better scores and may fail to note information regarding costs. An alternative approach is that used by the M-Health Index & Navigation Database, which presents each app characteristic or feature as a separate filter [44]. This is beneficial in that it allows consumers to decide which characteristics or features matter to them but is challenging as multiple fields and filters exist. A better understanding of what matters to consumers provides useful information to guide decisions regarding which information to provide and to improve systems providing information to consumers. Apps should be transparent about their costs, rather than hiding features behind a paywall, where it is not possible to evaluate the usefulness of those features before making a payment. This was particularly emphasized in this study, as we had to remove one of the apps from our study as it required users to input credit card details, which would automatically charge an annual subscription, despite offering a free trial. This is a consideration highlighted through the valuable input from the Young Person's Mental Health Advisory Group [22], reinforcing the importance of patient and public involvement.

### Implications

This study has significant implications for the use of mental health and well-being apps. We show that professionals' and app store reviews are insufficient for mental health app users to make informed decisions based on the aspects of apps that are important to them. This is even more important in the context of the COVID-19 pandemic, with disrupted usual mental health care and patients relying on web-based mental health support. The study findings suggest that additional work should be conducted to ensure mental health apps are as useful as possible in supporting the public's mental health. In addition, review platforms should seek to incorporate the views of those with mental health problems when publishing reviews to maximize their relevance for those most likely to use mental health apps.

### Strengths and Limitations

Existing research on what people think about health or mental health apps has focused on the perspectives of predominantly the White population (84%) [45]. Digital tools can help bridge inequalities in access to mental health care; therefore, it is essential to consider the perspectives of typically underserved



people. We improved on a prior work with a much more heterogeneous sample (only 28% of our sample was White British). Our sample was skewed toward women (15/21, 71%), but as women are more likely to use mental health apps [46] and the internet for health-related information [47], our sample may be representative of mental health app users in this respect. However, our sample was generally highly educated (13/21, 62%) to degree level. Although smartphone ownership is associated with higher levels of education [48], it is likely that mental health app use and opinions differ based on education, which we were unable to capture in our sample. Future work should aim to investigate differences in reviews of mental health apps with a larger and more diverse sample in terms of gender and educational attainment.

While we investigated whether the participants had prior experience using well-being apps, we did not directly measure whether they had previously used the same apps they used in the study. This may have affected the study; however, as we did not alter the apps in any way, their prior experience may simply corroborate our findings. It is also worth noting that although we refer to these apps as “mental health apps,” the included apps are all general wellness or well-being apps and not digital therapeutics. Distinctions between categories of apps to support the mental health and well-being of consumers are starting to emerge but are still murky as regulations and guidance attempt to catch up with this market. Future work can explore different groups’ understanding of these distinctions to understand what is acceptable for these low-intensity intervention apps.

This study was designed to understand the views of people with experience of mental health problems and so reflects the views

of those who are most likely to benefit from the support provided by mental health apps. Our participants had mental health problems, but future work should capture the opinions of a group of people with varying psychiatric diagnoses to understand whether those factors affect app ratings. For example, a randomized controlled trial found that using mental health apps was associated with improvements in depressive symptoms, but they had no effect on anxiety, compared with a control group not using a mental health app [49]. However, most studies found that mental health apps are effective in improving symptoms of both depression and anxiety [1,2,50,51], as well as the quality of sleep [52]. Studies on other conditions, including serious mental illnesses, are limited; therefore, future work should investigate differences in efficacy depending on symptomatology. If differences exist, then there may be differing priorities in mental health apps depending on their diagnoses.

## Conclusions

We found that participants with lived experiences of mental health problems rate apps differently than professionals and that these ratings correlate poorly with those publicly available on app stores. This is particularly important in the current climate of the COVID-19 pandemic, with more people seeking their mental health care on the web. Further research is needed to explore the perspectives of a diverse group of mental health service users. Our participants also emphasized aspects that are not currently captured in the available review systems. Our study findings suggest that aspects such as ease of use, engaging features and designs, low cost, and some educational content should be added in the future.

---

## Acknowledgments

The authors acknowledge the expert input of the Maudsley Biomedical Research Centre’s FAST-R group and Young Person’s Mental Health Advisory Group. This paper represents independent research funded by the National Institute for Health Research Biomedical Research Centre in South London and Maudsley National Health Service Foundation Trust and King’s College London (IS-BRC-1215-20018). Authors MN and SMS are funded by One Mind for the operation of One Mind PsyberGuide.

---

## Data Availability

The data that support the findings of this study can be obtained from the corresponding author, SJ, on reasonable request.

---

## Conflicts of Interest

SMS serves on the Scientific Advisory Board for Headspace, for which he receives compensation, and has also received consulting payments from Trusst (K Health) and Otsuka Pharmaceuticals.

---

### Multimedia Appendix 1

Predefined protocol for converting qualitative text into the binary classification of participant experience.

[\[DOCX File, 24 KB - formative\\_v6i9e39813\\_app1.docx\]](#)

---

### Multimedia Appendix 2

Full breakdown of the number of negative agreements, positive agreements, participant negatives, and professional negatives, and their relative percentages, for all domains.

[\[DOCX File, 26 KB - formative\\_v6i9e39813\\_app2.docx\]](#)

## References

1. Firth J, Torous J, Nicholas J, Carney R, Prata A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry* 2017 Oct;16(3):287-298 [FREE Full text] [doi: [10.1002/wps.20472](https://doi.org/10.1002/wps.20472)] [Medline: [28941113](https://pubmed.ncbi.nlm.nih.gov/28941113/)]
2. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *J Affect Disord* 2017 Aug 15;218:15-22 [FREE Full text] [doi: [10.1016/j.jad.2017.04.046](https://doi.org/10.1016/j.jad.2017.04.046)] [Medline: [28456072](https://pubmed.ncbi.nlm.nih.gov/28456072/)]
3. 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](https://doi.org/10.1002/wps.20673)] [Medline: [31496095](https://pubmed.ncbi.nlm.nih.gov/31496095/)]
4. Lipschitz JM, Connolly SL, Miller CJ, Hogan TP, Simon SR, Burdick KE. Patient interest in mental health mobile app interventions: demographic and symptom-level differences. *J Affect Disord* 2020 Feb 15;263:216-220. [doi: [10.1016/j.jad.2019.11.083](https://doi.org/10.1016/j.jad.2019.11.083)] [Medline: [31818779](https://pubmed.ncbi.nlm.nih.gov/31818779/)]
5. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: evaluation of mental health app store quality claims. *NPJ Digit Med* 2019 Mar 22;2:18 [FREE Full text] [doi: [10.1038/s41746-019-0093-1](https://doi.org/10.1038/s41746-019-0093-1)] [Medline: [31304366](https://pubmed.ncbi.nlm.nih.gov/31304366/)]
6. Marshall JM, Dunstan DA, Bartik W. Clinical or gimmickal: the use and effectiveness of mobile mental health apps for treating anxiety and depression. *Aust N Z J Psychiatry* 2020 Jan;54(1):20-28. [doi: [10.1177/0004867419876700](https://doi.org/10.1177/0004867419876700)] [Medline: [31552747](https://pubmed.ncbi.nlm.nih.gov/31552747/)]
7. Leigh S, Flatt S. App-based psychological interventions: friend or foe? *Evid Based Ment Health* 2015 Nov;18(4):97-99. [doi: [10.1136/eb-2015-102203](https://doi.org/10.1136/eb-2015-102203)] [Medline: [26459466](https://pubmed.ncbi.nlm.nih.gov/26459466/)]
8. Neary M, Schueller SM. State of the field of mental health apps. *Cogn Behav Pract* 2018 Nov;25(4):531-537 [FREE Full text] [doi: [10.1016/j.cbpra.2018.01.002](https://doi.org/10.1016/j.cbpra.2018.01.002)] [Medline: [33100810](https://pubmed.ncbi.nlm.nih.gov/33100810/)]
9. Torous J, Powell AC. Current research and trends in the use of smartphone applications for mood disorders. *Internet Interv* 2015 May;2(2):169-173. [doi: [10.1016/j.invent.2015.03.002](https://doi.org/10.1016/j.invent.2015.03.002)]
10. Leigh S. Comparing apples and oranges: barriers to evidence-based practice for app-based psychological interventions. *Evid Based Ment Health* 2016 Aug;19(3):90-92. [doi: [10.1136/eb-2016-102384](https://doi.org/10.1136/eb-2016-102384)] [Medline: [27431656](https://pubmed.ncbi.nlm.nih.gov/27431656/)]
11. Wykes T, Schueller S. Why reviewing apps is not enough: transparency for trust (T4T) principles of responsible health app marketplaces. *J Med Internet Res* 2019 May 02;21(5):e12390 [FREE Full text] [doi: [10.2196/12390](https://doi.org/10.2196/12390)] [Medline: [31045497](https://pubmed.ncbi.nlm.nih.gov/31045497/)]
12. Powell AC, Landman AB, Bates DW. In search of a few good apps. *JAMA* 2014 May 14;311(18):1851-1852. [doi: [10.1001/jama.2014.2564](https://doi.org/10.1001/jama.2014.2564)] [Medline: [24664278](https://pubmed.ncbi.nlm.nih.gov/24664278/)]
13. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015 Mar 11;3(1):e27 [FREE Full text] [doi: [10.2196/mhealth.3422](https://doi.org/10.2196/mhealth.3422)] [Medline: [25760773](https://pubmed.ncbi.nlm.nih.gov/25760773/)]
14. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-mental health apps in Canada: results of a modified Delphi process. *JMIR Mhealth Uhealth* 2018 Jul 09;6(7):e10016 [FREE Full text] [doi: [10.2196/10016](https://doi.org/10.2196/10016)] [Medline: [29986846](https://pubmed.ncbi.nlm.nih.gov/29986846/)]
15. ORCHA. ORCHA: Unlocking the power of Digital Health for the population. Organisation for the Review of Health and Care Apps. 2022. URL: <https://orchahealth.com/> [accessed 2022-02-22]
16. Martens D, Maalej W. Towards understanding and detecting fake reviews in app stores. *Empir Software Eng* 2019 May 10;24(6):3316-3355. [doi: [10.1007/s10664-019-09706-9](https://doi.org/10.1007/s10664-019-09706-9)]
17. Niedzwiedz CL, Green MJ, Benzeval M, Campbell D, Craig P, Demou E, et al. Mental health and health behaviours before and during the initial phase of the COVID-19 lockdown: longitudinal analyses of the UK Household Longitudinal Study. *J Epidemiol Community Health* 2021 Mar;75(3):224-231 [FREE Full text] [doi: [10.1136/jech-2020-215060](https://doi.org/10.1136/jech-2020-215060)] [Medline: [32978210](https://pubmed.ncbi.nlm.nih.gov/32978210/)]
18. Pieh C, Budimir S, Delgadillo J, Barkham M, Fontaine JR, Probst T. Mental health during COVID-19 lockdown in the United Kingdom. *Psychosom Med* 2021 May 01;83(4):328-337. [doi: [10.1097/PSY.0000000000000871](https://doi.org/10.1097/PSY.0000000000000871)] [Medline: [33009276](https://pubmed.ncbi.nlm.nih.gov/33009276/)]
19. Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. *Lancet Psychiatry* 2020 Oct;7(10):883-892 [FREE Full text] [doi: [10.1016/S2215-0366\(20\)30308-4](https://doi.org/10.1016/S2215-0366(20)30308-4)] [Medline: [32707037](https://pubmed.ncbi.nlm.nih.gov/32707037/)]
20. On World Mental Health Day, ORCHA reports that the use of mental health apps is still on the rise. Organisation for the Review of Health and Care Apps. 2021 Oct 8. URL: <https://www.orchahealth.com/on-world-mental-health-day-orcha-reports-that-the-use-of-mental-health-apps-is-still-on-the-rise/> [accessed 2022-02-22]
21. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. *J Med Internet Res* 2018 Jun 11;20(6):e10141 [FREE Full text] [doi: [10.2196/10141](https://doi.org/10.2196/10141)] [Medline: [29891468](https://pubmed.ncbi.nlm.nih.gov/29891468/)]
22. Hudson G. Young Person's Mental Health Advisory Group. King's College London. 2022. URL: <https://www.kcl.ac.uk/research/ypmhag> [accessed 2022-01-26]

23. Glick G, Druss B, Pina J, Lally C, Conde M. Use of mobile technology in a community mental health setting. *J Telemed Telecare* 2016 Oct;22(7):430-435. [doi: [10.1177/1357633X15613236](https://doi.org/10.1177/1357633X15613236)] [Medline: [26519378](https://pubmed.ncbi.nlm.nih.gov/26519378/)]
24. Abu Rahal Z, Vadas L, Manor I, Bloch B, Avital A. Use of information and communication technologies among individuals with and without serious mental illness. *Psychiatry Res* 2018 Aug;266:160-167. [doi: [10.1016/j.psychres.2018.05.026](https://doi.org/10.1016/j.psychres.2018.05.026)] [Medline: [29864616](https://pubmed.ncbi.nlm.nih.gov/29864616/)]
25. Neary M, Bunyi J, Palomares K, Mohr DC, Powell A, Ruzek J, et al. A process for reviewing mental health apps: using the one mind PsyberGuide credibility rating system. *Digit Health* 2021 Oct 29;7:20552076211053690 [FREE Full text] [doi: [10.1177/20552076211053690](https://doi.org/10.1177/20552076211053690)] [Medline: [34733541](https://pubmed.ncbi.nlm.nih.gov/34733541/)]
26. Lagan S, D'Mello R, Vaidyam A, Bilden R, Torous J. Assessing mental health apps marketplaces with objective metrics from 29,190 data points from 278 apps. *Acta Psychiatr Scand* 2021 Aug;144(2):201-210. [doi: [10.1111/acps.13306](https://doi.org/10.1111/acps.13306)] [Medline: [33835483](https://pubmed.ncbi.nlm.nih.gov/33835483/)]
27. O'Loughlin K, Neary M, Adkins EC, Schueller SM. Reviewing the data security and privacy policies of mobile apps for depression. *Internet Interv* 2019 Mar;15:110-115 [FREE Full text] [doi: [10.1016/j.invent.2018.12.001](https://doi.org/10.1016/j.invent.2018.12.001)] [Medline: [30792962](https://pubmed.ncbi.nlm.nih.gov/30792962/)]
28. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and Web-based eHealth interventions. *J Med Internet Res* 2017 Mar 21;19(3):e82 [FREE Full text] [doi: [10.2196/jmir.7270](https://doi.org/10.2196/jmir.7270)] [Medline: [28325712](https://pubmed.ncbi.nlm.nih.gov/28325712/)]
29. 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](https://doi.org/10.2196/14567)] [Medline: [31573916](https://pubmed.ncbi.nlm.nih.gov/31573916/)]
30. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960 Apr 1;20(1):37-46. [doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)]
31. IBM SPSS Statistics for Windows, Version 27.0. IBM Corp. 2020. URL: <https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-27> [accessed 2022-03-17]
32. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
33. Jilka S, Hudson G, Jansli SM, Negbenose E, Wilson E, Odoi CM, et al. How to make study documents clear and relevant: the impact of patient involvement. *BJPsych open* 2021 Nov 04;7(6):E202. [doi: [10.1192/bjo.2021.1040](https://doi.org/10.1192/bjo.2021.1040)]
34. Hudson G, Jansli SM, Erturk S, Morris D, Odoi CM, Clayton-Turner A, et al. Investigation of carers' perspectives of dementia misconceptions on Twitter: focus group study. *JMIR Aging* 2022 Jan 24;5(1):e30388 [FREE Full text] [doi: [10.2196/30388](https://doi.org/10.2196/30388)] [Medline: [35072637](https://pubmed.ncbi.nlm.nih.gov/35072637/)]
35. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. *BMJ* 2000 Jan 08;320(7227):114-116 [FREE Full text] [doi: [10.1136/bmj.320.7227.114](https://doi.org/10.1136/bmj.320.7227.114)] [Medline: [10625273](https://pubmed.ncbi.nlm.nih.gov/10625273/)]
36. Sweeney A, Greenwood KE, Williams S, Wykes T, Rose DS. Hearing the voices of service user researchers in collaborative qualitative data analysis: the case for multiple coding. *Health Expect* 2013 Dec;16(4):e89-e99 [FREE Full text] [doi: [10.1111/j.1369-7625.2012.00810.x](https://doi.org/10.1111/j.1369-7625.2012.00810.x)] [Medline: [22958162](https://pubmed.ncbi.nlm.nih.gov/22958162/)]
37. NVivo (Version 12). QSR International. 2018. URL: [https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home?\\_ga=2.159471485.1135045820.1647520139-738635854.1647520139](https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home?_ga=2.159471485.1135045820.1647520139-738635854.1647520139) [accessed 2022-03-17]
38. Bry LJ, Chou T, Miguel E, Comer JS. Consumer smartphone apps marketed for child and adolescent anxiety: a systematic review and content analysis. *Behav Ther* 2018 Mar;49(2):249-261 [FREE Full text] [doi: [10.1016/j.beth.2017.07.008](https://doi.org/10.1016/j.beth.2017.07.008)] [Medline: [29530263](https://pubmed.ncbi.nlm.nih.gov/29530263/)]
39. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela JP, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. *Evid Based Ment Health* 2019 Feb;22(1):4-9 [FREE Full text] [doi: [10.1136/ebmental-2018-300069](https://doi.org/10.1136/ebmental-2018-300069)] [Medline: [30635262](https://pubmed.ncbi.nlm.nih.gov/30635262/)]
40. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. *J Med Internet Res* 2017 May 25;19(5):e176 [FREE Full text] [doi: [10.2196/jmir.7332](https://doi.org/10.2196/jmir.7332)] [Medline: [28546138](https://pubmed.ncbi.nlm.nih.gov/28546138/)]
41. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. *J Med Internet Res* 2021 Jan 13;23(1):e17828 [FREE Full text] [doi: [10.2196/17828](https://doi.org/10.2196/17828)] [Medline: [33439133](https://pubmed.ncbi.nlm.nih.gov/33439133/)]
42. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. *J Gen Intern Med* 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: [10.1007/s11606-016-3771-6](https://doi.org/10.1007/s11606-016-3771-6)] [Medline: [27418347](https://pubmed.ncbi.nlm.nih.gov/27418347/)]
43. Thornton LK, Kay-Lambkin FJ. Specific features of current and emerging mobile health apps: user views among people with and without mental health problems. *Mhealth* 2018 Dec 6;4:56 [FREE Full text] [doi: [10.21037/mhealth.2018.11.04](https://doi.org/10.21037/mhealth.2018.11.04)] [Medline: [30701174](https://pubmed.ncbi.nlm.nih.gov/30701174/)]
44. Beth Israel Deaconess Medical Center. 2020. URL: <https://mindapps.org/> [accessed 2022-05-16]
45. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. *J Med Internet Res* 2013 Apr 18;15(4):e86 [FREE Full text] [doi: [10.2196/jmir.2583](https://doi.org/10.2196/jmir.2583)] [Medline: [23598614](https://pubmed.ncbi.nlm.nih.gov/23598614/)]

46. Antezana G, Venning A, Smith D, Bidargaddi N. Do young men and women differ in well-being apps usage? Findings from a randomised trial. *Health Informatics J* 2022;28(1):14604582211064825 [FREE Full text] [doi: [10.1177/14604582211064825](https://doi.org/10.1177/14604582211064825)] [Medline: [35128952](https://pubmed.ncbi.nlm.nih.gov/35128952/)]
47. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related Internet usage and design feature preference for e-mental health programs among men and women. *J Med Internet Res* 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: [10.2196/11224](https://doi.org/10.2196/11224)] [Medline: [30882361](https://pubmed.ncbi.nlm.nih.gov/30882361/)]
48. Mobile Fact Sheet. Pew Research Center. 2021 Apr 7. URL: <http://www.pewinternet.org/fact-sheet/mobile/> [accessed 2022-03-24]
49. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. *Behav Res Ther* 2018 Oct;109:75-83. [doi: [10.1016/j.brat.2018.08.003](https://doi.org/10.1016/j.brat.2018.08.003)] [Medline: [30125790](https://pubmed.ncbi.nlm.nih.gov/30125790/)]
50. Mohr DC, Kwasny MJ, Meyerhoff J, Graham AK, Lattie EG. The effect of depression and anxiety symptom severity on clinical outcomes and app use in digital mental health treatments: meta-regression of three trials. *Behav Res Ther* 2021 Dec;147:103972. [doi: [10.1016/j.brat.2021.103972](https://doi.org/10.1016/j.brat.2021.103972)] [Medline: [34600398](https://pubmed.ncbi.nlm.nih.gov/34600398/)]
51. Planas R, Yuguero O. Technological prescription: evaluation of the effectiveness of mobile applications to improve depression and anxiety. Systematic review. *Inform Health Soc Care* 2021 Sep 02;46(3):273-290. [doi: [10.1080/17538157.2021.1887196](https://doi.org/10.1080/17538157.2021.1887196)] [Medline: [33685325](https://pubmed.ncbi.nlm.nih.gov/33685325/)]
52. Huberty J, Puzia ME, Larkey L, Vranceanu AM, Irwin MR. Can a meditation app help my sleep? A cross-sectional survey of Calm users. *PLoS One* 2021 Oct 22;16(10):e0257518 [FREE Full text] [doi: [10.1371/journal.pone.0257518](https://doi.org/10.1371/journal.pone.0257518)] [Medline: [34679078](https://pubmed.ncbi.nlm.nih.gov/34679078/)]

## Abbreviations

**MARS:** Mobile App Rating Scale

*Edited by A Mavragani; submitted 27.05.22; peer-reviewed by A Peipert, JD Bacsu; comments to author 22.06.22; revised version received 29.07.22; accepted 09.08.22; published 23.09.22.*

*Please cite as:*

*Hudson G, Negbenose E, Neary M, Jansli SM, Schueller SM, Wykes T, Jilka S*

*Comparing Professional and Consumer Ratings of Mental Health Apps: Mixed Methods Study*

*JMIR Form Res* 2022;6(9):e39813

URL: <https://formative.jmir.org/2022/9/e39813>

doi: [10.2196/39813](https://doi.org/10.2196/39813)

PMID: [36149733](https://pubmed.ncbi.nlm.nih.gov/36149733/)

©Georgie Hudson, Esther Negbenose, Martha Neary, Sonja M Jansli, Stephen M Schueller, Til Wykes, Sagar Jilka. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 23.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Using Social Media to Facilitate Communication About Women's Testing: Tool Validation Study

Tara Coffin<sup>1</sup>, BA, MEd, PhD; Deborah Bowen<sup>1</sup>, PhD; Karen Lu<sup>2</sup>, MD; Elizabeth M Swisher<sup>1</sup>, MD; Nadine Rayes<sup>2</sup>, MS; Barbara Norquist<sup>1</sup>, MD; Stephanie V Blank<sup>3</sup>, MD; Douglas A Levine<sup>4</sup>, MD; Jamie Nadine Bakkum-Gamez<sup>5</sup>, MD; Gini F Fleming<sup>6</sup>, MD; Olufunmilayo I Olopade<sup>6</sup>, MD; Iris Romero<sup>6</sup>, MD; Alan D'Andrea<sup>7</sup>, MD; Denise R Nebgen<sup>2</sup>, MD, PhD; Christine Peterson<sup>2</sup>, PhD; Mark F Munsell<sup>2</sup>, MS; Kathleen Gavin<sup>8</sup>, MPH; Jamie Crase<sup>1</sup>; Deborah Polinsky<sup>9</sup>, MS; Rebecca Lechner<sup>8</sup>

<sup>1</sup>University of Washington, Seattle, WA, United States

<sup>2</sup>MD Anderson Cancer Center, Houston, TX, United States

<sup>3</sup>Mount Sinai Medical Center, New York, NY, United States

<sup>4</sup>New York University Langone Health, New York, NY, United States

<sup>5</sup>Mayo Clinic, Rochester, MN, United States

<sup>6</sup>University of Chicago, Chicago, IL, United States

<sup>7</sup>Dana-Farber Cancer Institute, Boston, MA, United States

<sup>8</sup>Minnesota Ovarian Cancer Alliance, Minneapolis, MN, United States

<sup>9</sup>SHARE, New York, NY, United States

**Corresponding Author:**

Tara Coffin, BA, MEd, PhD

University of Washington

1400 NE Campus Parkway

Seattle, WA, 98195-4550

United States

Phone: 1 (206) 543 2100

Email: [tbcoffin@uw.edu](mailto:tbcoffin@uw.edu)

## Abstract

**Background:** Strong participant recruitment practices are critical to public health research but are difficult to achieve. Traditional recruitment practices are often time consuming, costly, and fail to adequately target difficult-to-reach populations. Social media platforms such as Facebook are well-positioned to address this area of need, enabling researchers to leverage existing social networks and deliver targeted information. The MAGENTA (Making Genetic Testing Accessible) study aimed to improve the availability of genetic testing for hereditary cancer susceptibility in at-risk individuals through the use of a web-based communication system along with social media advertisements to improve reach.

**Objective:** This paper is aimed to evaluate the effectiveness of Facebook as an outreach tool for targeting women aged  $\geq 30$  years for recruitment in the MAGENTA study.

**Methods:** We designed and implemented paid and unpaid social media posts with ongoing assessment as a primary means of research participant recruitment in collaboration with patient advocates. Facebook analytics were used to assess the effectiveness of paid and unpaid outreach efforts.

**Results:** Over the course of the reported recruitment period, Facebook materials had a reach of 407,769 people and 57,248 (14.04%) instances of engagement, indicating that approximately 14.04% of people who saw information about the study on Facebook engaged with the content. Paid advertisements had a total reach of 373,682. Among those reached, just <15% (54,117/373,682, 14.48%) engaged with the page content. Unpaid posts published on the MAGENTA Facebook page resulted in a total of 34,087 reach and 3131 instances of engagement, indicating that around 9.19% (3131/34,087) of people who saw unpaid posts engaged. Women aged  $\geq 65$  years reported the best response rate, with approximately 43.95% (15,124/34,410) of reaches translating to engagement. Among the participants who completed the eligibility questionnaire, 27.44% (3837/13,983) had heard about the study through social media or another webpage.

**Conclusions:** Facebook is a useful way of enhancing clinical trial recruitment of women aged  $\geq 30$  years who have a potentially increased risk for ovarian cancer by promoting news stories over social media, collaborating with patient advocacy groups, and running paid and unpaid campaigns.

**Trial Registration:** ClinicalTrials.gov NCT02993068; <https://clinicaltrials.gov/ct2/show/NCT02993068>

(*JMIR Form Res* 2022;6(9):e35035) doi:[10.2196/35035](https://doi.org/10.2196/35035)

## KEYWORDS

ovarian cancer; hereditary cancer; genetic testing; online social media recruitment; Facebook; social media; mobile phone

## Introduction

### Background

High participant response rates and recruitment yields are critical to public health research but are difficult to achieve [1-3]. Traditional recruitment practices, including radio or newspaper advertising, in-person referrals, and flyers, are often time consuming to implement, costly, and fail to adequately target difficult-to-reach populations [4,5]. The initial net cast using these types of recruitment methods may result in a high number of interested parties; however, such efforts result in proportionately fewer eligible and enrolled participants, and certain demographics are frequently left underrepresented [6]. Social media is well-positioned to address many of these issues and improve participant recruitment by providing new platforms for people to learn about public health research [7-10].

The term *social media* broadly describes a variety of web-based social networking platforms or web-based spaces where the public can generate, engage with, and share information, including platforms such as Facebook, Twitter, and Instagram [11,12]. Social media enables researchers to deliver information to a wide audience; target specific groups of people, including hard-to-reach subpopulations; and adapt outreach efforts on an ongoing basis [7-10]. Current research indicates that social media recruitment methods are an improvement over traditional methods in terms of both cost and effectiveness [13-16].

Facebook, used by more than three-quarters of adults on the web, is particularly well-suited for research recruitment [17,18]. Over Facebook, users can engage with user-generated content, publish photos on their Facebook pages, post status updates, and share information with friends and family. Users follow the content of interest and engage socially with paid advertisements and other content. Researchers can leverage this environment, creating content tailored for specific populations using online behavioral advertising (OBA) and respondent-driven sampling to improve reach [19].

OBA data can help researchers improve their marketing reach. OBA data include information collected from a broad range of web-based sources about the behaviors that users exhibit on the web [20]. OBA appeals to researchers in public health, seeking to improve recruitment tools and offering an alternative outreach method with a broader reach that may overcome certain recruitment barriers, such as geographic limitations [21-25]. Instead of wondering whether a flyer is posted in the right place for the right type of individual to see, researchers can guarantee that their message is being displayed to the intended person. This approach is not without its limitations, and some health

professionals and researchers have expressed reluctance, citing concerns about biased sampling or reach that may accompany social media platforms [26] and privacy [12,27,28].

Facebook also allows public health professionals to leverage existing social networks through snowball sampling [29,30]. Snowball sampling, which has traditionally taken place offline, can capitalize on existing web-based social networks, such as patient advocacy groups [30,31]. By encouraging a small sample of a target population to refer others to a research study, snowball sampling helps researchers access hidden subpopulations that are typically difficult to sample using traditional recruitment methods [30]. From snowball sampling to inviting opportunities to shape the tone, imagery, and content to fit the needs of the intended audience, social media is well-positioned to function as a targeted communication tool. With these advantages, social media has the potential to take traditional snowball sampling one step further, enabling researchers to potentially connect with harder-to-reach populations [32]. This quality grants social media recruitment the ability to potentially shift the pattern of health inequities, improving the representation of certain communities in the research arena [33].

Recent reviews indicate that most studies using Facebook to recruit participants for health research have focused on people aged 18 to 30 years [8,34]. In comparison, few studies have evaluated social media as a means of recruiting people affected by cancer who are aged  $\geq 35$  years [34], and no studies have explored how social media recruitment performs when targeting women at risk for ovarian cancer. The consensus is that older people may be less likely to adopt new technologies, such as social media [34,35]. Other studies have reported high reach but low engagement among social media users, resulting in a high attrition rate for social media recruitment [36]. However, this research failed to examine advertisement content or take the growth of the social media platform into consideration. As the social media base continues to grow, the profile of the average user evolves, and with it, the age of the average Facebook user continues to increase [37]. With this evolution in mind, ongoing assessment is needed to evaluate the effectiveness of social media for research participant recruitment across different demographics, and more research is needed to better understand how Facebook functions as a recruitment tool in the context of ovarian cancer [20].

### Study Aims

This research sought to determine whether Facebook is an effective recruitment tool for targeting women aged  $\geq 30$  years for recruitment into the MAGENTA (Making Genetic Testing

Accessible) study by evaluating innovative methods for the recruitment of research participants using Facebook. To accomplish this objective, a series of posts and advertisements, including paid and unpaid posts, were published and assessed on an ongoing basis. These materials used a variety of imagery and languages and leveraged Facebook's OBA tools to target specific populations and eligible participants. We hypothesized that unpaid Facebook posts and Facebook advertisements would improve the reach of the study material and result in improved study enrollment.

## Methods

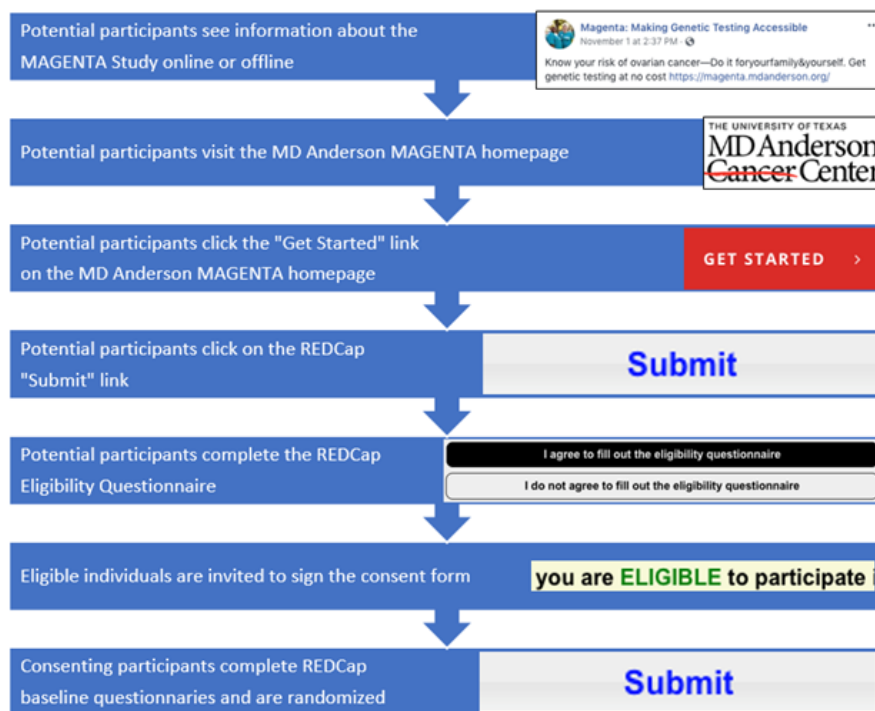
### About the MAGENTA Study

The MAGENTA study was a nationwide Stand Up To Cancer initiative that sought to improve access to genetic testing for ovarian cancer. The study recruited and randomized 3839 women from the United States with a potentially increased risk of ovarian cancer. Participants were randomized to 1 of 4 arms, receiving a combination of pretest or posttest telephone genetic counseling and pretest or posttest web-based education with optional telephone counseling [38]. The active recruitment period took place between April 2017 and January 2020. This

study received institutional review board approval from the MD Anderson Cancer Center and was a collaborative effort that included several cancer research centers and patient advocacy groups, including the Ovarian Cancer Research Alliance, National Ovarian Cancer Coalition, and Minnesota Ovarian Cancer Alliance.

Once potential participants had learned about the MAGENTA study, they were prompted to visit the study website. From there, interested parties clicked to participate in the web-based communication system, starting with study information and then moving through the eligibility screen, informed consent, and enrollment (Figure 1). Data were collected at baseline and follow-up using REDCap (Research Electronic Data Capture), an electronic survey tool sponsored by the University of Washington (WA). All outreach materials received institutional review board review through the MD Anderson Cancer Center. The results of the MAGENTA study indicate that electronic genetic education and results released without genetic counseling were noninferior with regard to patient distress. Importantly, research also found that providing genetic education and results in this capacity was associated with higher test completion and lower distress [38].

**Figure 1.** Illustration of the web-based communication system used by the MAGENTA (Making Genetic Testing Accessible) study. REDCap: Research Electronic Data Capture.



### Developing a Media Kit

Adapting the methods outlined by Carter-Harris et al [39] and Musiat et al [40], the study media kit was developed in collaboration with key stakeholders. This group comprised health care professionals from cancer care and research centers and patient advocates from advocacy groups across the United States, including those listed previously. Patient advocates were consulted extensively during the development of the study materials, including the media kit described in the following

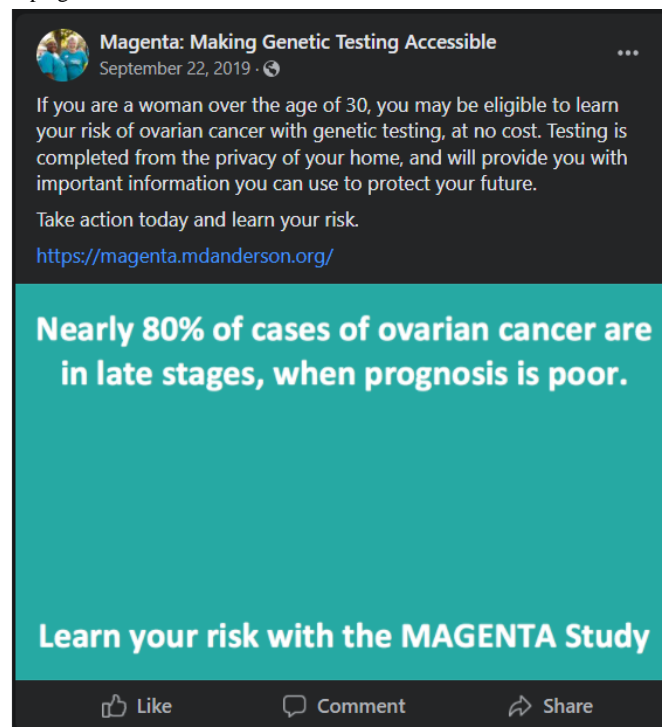
sections. The media kit included Facebook recruitment materials and a list of social media contacts, such as patient advocacy groups and other groups with an apparent interest in breast and ovarian cancer.

The media kit also included different types of posts generated for recruitment purposes, including paid advertisements, unpaid posts, sample tweets, a list of relevant hashtags to incorporate into posts, and a selection of media for use across all social media posts and advertisements (example posts can be reviewed in Figures 2-4). Unpaid Facebook posts and paid advertisements

included at least one media component, a brief description of the study, relevant hashtags, and a link to the study home page (Figure 1). A MAGENTA Facebook page was created to develop trust with potential participants [41,42]. The Facebook page provided basic information about the study, served as a platform for sharing unpaid and paid social media posts, and directed potential participants to the study website. Materials from the media kit were assessed by patient advocates and underwent usability testing. Advertisements and posts were created with tone and imagery in mind, focusing on content related to ovarian cancer research that elicited a combination of the following concepts adapted from Batterham [43]:

1. Content instills a sense of collaboration, conveying the idea that one is participating in research as a member of a team to address a health problem (in this case, ovarian cancer was framed as the problem).
2. Content instills a sense of independence or conveys the idea that one is addressing the problem of ovarian cancer as an individual through research participation.
3. Content instills a sense of altruism, conveying the idea that the individual is participating in research for the benefit of others.
4. Content instills a sense of self-gain or self-preservation, conveying the idea that the individual is participating in research for personal gain.

**Figure 2.** An example Facebook post containing a still image, study link, and brief description of the outreach. This type of post was used in both for unpaid posting and paid advertising campaigns.

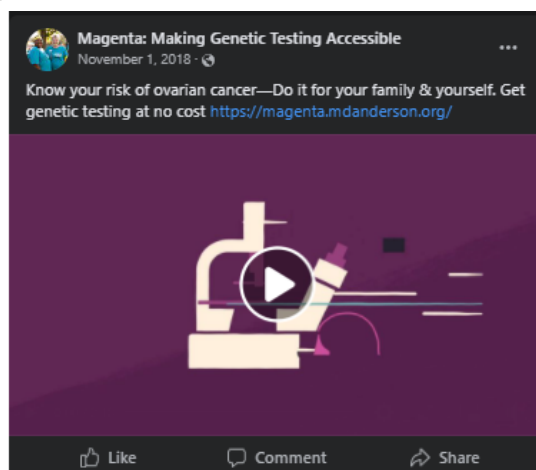


**Figure 3.** An example of a Facebook post sharing the WCCO news story, which includes a video of the news story and a brief text section. This type of post is an example of a boosted post that was used for unpaid posting.





**Figure 4.** An example of a Facebook post containing the study video, study link, and a brief description of the outreach. This type of post was used in both unpaid and paid advertising campaigns.



### Publishing Paid Advertisements and Unpaid Posts

Unpaid posts were published directly on the MAGENTA Facebook page on a regular basis and on patient advocacy Facebook pages. Paid advertisements were published using Facebook's advertising tool. Once the objective or goal of the campaign (eg, post engagement, website clicks, and video views) was set, the audience was identified using Facebook's audience-targeting tool. Targeted populations for the purposes of this study included English-speaking women  $\geq$ aged 30 years living in the United States. Additional geographic and behavioral targeting was included on a case-by-case basis and is described in greater detail in [Table 1](#).

Census data were used to inform additional geographic and socioeconomic targeting and included data surrounding

racial-ethnic groups and the rurality of the location. These variables were layered using ArcGIS Pro (version 2.5; Esri) to select the specific geographic targets. ArcGIS is a mapping and analysis tool that allows users to use a geographic information system to capture, manipulate, and analyze geospatial data. Once the audience was selected, advertising content was uploaded to Facebook, a campaign budget was selected, and a campaign schedule was set. On the basis of the intended audience, Facebook uses OBA approaches to push out content with the above parameters in mind. Although targeting affects results, over social media platforms, including Facebook, the budget arguably has the most impact on reach, and larger budgets are generally associated with more results, assuming that appropriate targeting is used.

**Table 1.** Description of Facebook paid campaign content and audience.

Advertisement name	Advertisement description			Audience description	
	Objective	Media	Age (years)	Geographic	Interests or additional targeting
Lookalike Audience Campaign	Conversions	Video	30 to ≥65	United States	Lookalike (United States, 1%)—website traffic
CA <sup>a</sup> Campaign 1	Conversions	Image	30 to 55	Salinas, CA	Breast cancer awareness, National Breast Cancer Foundation, and National Ovarian Cancer Coalition
CA Campaign 2	Conversions	Video	30 to ≥65	Salinas, CA	Advocacy groups for patients with multiple hereditary cancers, focusing on ovarian and breast cancer
CA Campaign 3	Conversions	Image	30 to 60	Downey, Pomona, Santa Ana, CA	Breast cancer awareness and Telemundo or Univision
Cascade Testing 1	Conversions	Image	30 to 55	United States	Facing Our Risk of Cancer Empowered
Cascade Testing 2	Conversions	Image	30 to 55	United States	Facing Our Risk of Cancer Empowered
African American Target Campaign	Traffic	Video	30 to ≥65	Birmingham, Montgomery, Alabama; Miami Gardens, Florida; Savannah, Georgia; New Orleans, Los Angeles; Baltimore, Maryland; Detroit, Flint, MI; Jackson, Mississippi; Memphis, Tennessee	N/A <sup>b</sup>
NY <sup>c</sup> Campaign	Traffic	Video	30 to ≥65	New York state	Advocacy groups for patients with multiple hereditary cancers, focusing on ovarian and breast cancer
BCOC <sup>d</sup> Interest Campaign	Traffic	Image	30 to ≥65	United States	N/A
BCOC Interest Campaign	Engagement	Image	30 to ≥65	Seattle, WA <sup>e</sup>	The Breast Cancer Research Foundation, national breast cancer Awareness Month, Susan G Komen for the Cure or Living Beyond Breast Cancer
BCOC Interest Campaign	Traffic	Image	30 to ≥65	United States	N/A
BCOC Interest Campaign	Traffic	Study video	30 to 50	United States	Advocacy groups for patients with multiple hereditary cancers, focusing on ovarian and breast cancer; keywords related to breast and ovarian cancer
BCOC Interest Campaign	Traffic	Study video	30 to 50	United States	Advocacy groups for patients with multiple hereditary cancers, focusing on ovarian and breast cancer; keywords related to breast and ovarian cancer
BCOC Interest Campaign	Traffic	Study video	30 to 50	United States	Advocacy groups for patients with multiple hereditary cancers, focusing on ovarian and breast cancer; keywords related to breast and ovarian cancer
Latino Target Campaign	Traffic	Image	30 to ≥65	Brownsville, El Paso, Laredo, McAllen, Texas; Hialeah, Florida; Downey, Oxnard, Pomona, Salinas, Santa Ana, CA	N/A

Advertisement name	Advertisement description			Audience description	
	Objective	Media	Age (years)	Geographic	Interests or additional targeting
BCOC Interest Campaign	Traffic	Image	30 to ≥65	United States	Advocacy groups for patients with multiple hereditary cancers, focusing on ovarian and breast cancer; keywords related to breast and ovarian cancer
BCOC Interest Campaign	Traffic	Study video	30 to ≥65	United States	Advocacy groups for patients with multiple hereditary cancers, focusing on ovarian and breast cancer; keywords related to breast and ovarian cancer
WA Campaign	Conversions	Image	30 to 50	Seattle, WA	Essence (magazine) and Latina (magazine)

<sup>a</sup>CA: California.

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

<sup>c</sup>NY: New York.

<sup>d</sup>BCOC: breast cancer and ovarian cancer.

<sup>e</sup>WA: Washington.

## Other Recruitment Efforts

All Facebook recruitment efforts took place alongside traditional recruitment efforts as part of the MAGENTA study. Traditional recruitment efforts included clinician referrals, direct emails from patient advocacy groups, the dissemination of study information at provider and patient advocate conferences, and sharing physical flyers in patient advocacy and clinical settings. Traditional efforts were largely based on the participating cancer research institutes, organizations, and patient advocacy groups. Study recruitment commenced with traditional methods, allowing for a controlled launch that allowed for an additional real-time usability assessment of the web-based communication system. In this first round of recruitment, enrollment relied primarily on word of mouth and flyer dissemination, both of which were facilitated by collaborating with patient advocacy groups. Following this controlled outreach, the study team expanded the outreach to include social media posts, as described above, in an effort to expand the reach of messaging.

## Evaluating Paid Advertisements and Unpaid Posts

Facebook analytics captured how users interacted with the social media MAGENTA content. Analytics included, among others, the following: *engagement* is defined as any time an individual takes action on the post, where action includes a click, comment, share, or view; *results* are defined as the number of times an advertisement achieved a specific outcome, delineated by the campaign objective; *reach* is defined as the number of people who saw the advertisement at least once; *impressions* are the number of times an advertisement was on a screen; *clicks* refer to the number of times someone clicked on the advertisement; and *video plays*. The study team also reviewed the *cost per result*. *Cost per result* was calculated by dividing the total amount of money spent by the number of *results*, which may include the number of video views or website visits, for example, obtained over the course of the campaign. Analytics,

including cost, were reviewed daily to assess the effectiveness and provide opportunities to adjust campaign content or targeting. The same information was collected for unpaid posts published directly on the MAGENTA study's Facebook page. If at any time the MAGENTA study website or another part of the web-based communication system became overburdened, advertisements were pulled, or turned off, until the traffic subsided.

## Ethics Approval

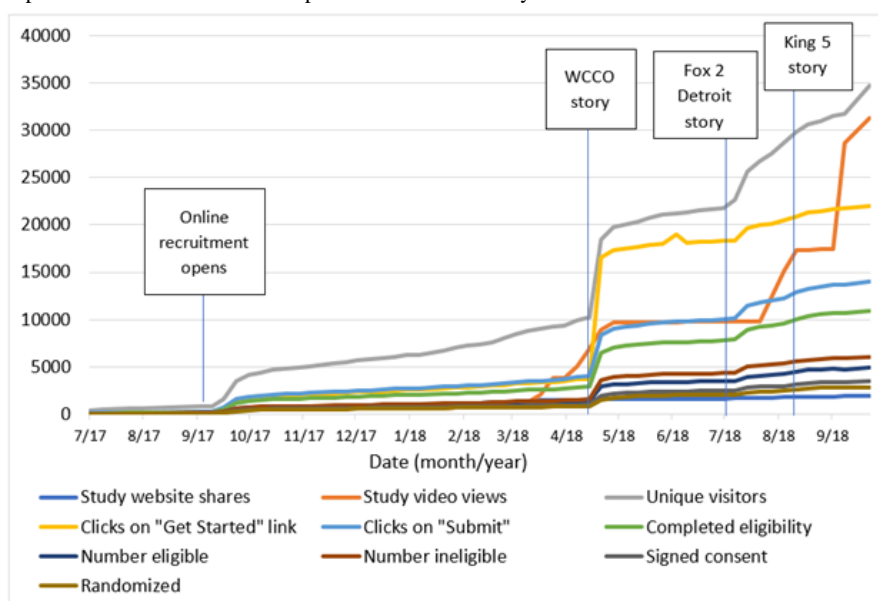
This study, including all outreach materials, received institutional review board review through MD Anderson Cancer Center (2016-0298).

## Results

### Overview

Active social media recruitment for the MAGENTA study took place between September 2017 and October 2018. The MAGENTA study relied on traditional recruitment methods starting in April 2017 until September 2017. Traditional recruitment methods continued throughout the social media recruitment period; however, the study team focused on web-based recruitment efforts in the interest of improving the reach across all 50 states. The recruitment timeline can be viewed in [Figure 5](#). During the active social media recruitment period, Facebook materials reached a total of 407,769 users, generating 57,248 (14.04%) instances of engagement, suggesting that approximately 14.04% of people who saw information about the MAGENTA study on Facebook engaged with the content. These numbers did not identify unique users and excluded posts published on Facebook pages managed by other breast and ovarian cancer groups. During this time, the MAGENTA study home page was shared 1948 times, and the MAGENTA study video was viewed 31,358 times ([Table 2](#)).

**Figure 5.** Timeline of enrollment trends and recruitment events captured during the active social media recruitment period (September 2017 to October 2018) and the number of responses received at different steps in recruitment activity.



**Table 2.** Summary of enrollment and randomization data.

Step in enrollment protocol	Total unique visitors
<b>MAGENTA<sup>a</sup> Facebook page content, n</b>	
Reach	407,769
Engagement (clicks, reactions, comments, and Facebook content shares)	57,248
Study website shares	1948
Study video views	31,358
<b>MAGENTA MD Anderson webpage</b>	
Unique visitors, N	34,715
Clicks on <i>Get Started</i> link, n (%)	22,029 (63.5)
<b>REDCap<sup>b</sup> introductory message, n (%)</b>	
Clicks on <i>Submit</i> link	14,025 (40.4)
<b>REDCap eligibility, n (%)</b>	
Started eligibility questionnaire	13,983 (40.3)
Completed eligibility questionnaire	10,883 (31.3)
Number eligible	4887 (14.1)
Number ineligible	5996 (17.3)

<sup>a</sup>MAGENTA: Making Genetic Testing Accessible.

<sup>b</sup>REDCap: Research Electronic Data Capture.

### Users Learn About the MAGENTA Study Over Television and Social Media

Of the 13,983 respondents to the MAGENTA REDCap eligibility questionnaire during the active social media recruitment period, <1% (n=23, 0.16%) indicated that they learned about the study from a magazine, <1% (n=86, 0.62%) from the radio, <2% (n=253, 1.81%) from a health care provider, >3% (n=459, 3.28%) from a patient advocacy group, and <8% (n=1102, 7.88%) from a friend. Approximately 8.64% (1209/13,983) indicated that they learned about the study from

a family member, whereas 27.44% (3837/13,983) indicated that they learned about the study on the web, either from social media or another webpage, and 28.16% (3938/13,983) from television. Among those who reported that they learned about the study from the internet, 16.94% (2369/13,983) specifically cited social media. A total of 21.7% (3034/13,983) of individuals who responded to the REDCap eligibility questionnaire did not indicate where they had heard about the study. Some respondents reported learning about the study from more than one source.

## Social Media Response

Paid advertisements (Table 3) reported a total reach of 373,682 during the active social media recruitment period. Among those reached, 3.57% (13,357/373,682) clicked on a link and 14.48% (54,117/373,682) engaged with the page content. Paid campaigns also generated 19,792 video plays and 9095 website conversions, which were defined as instances where a potential participant viewed the page content on the MAGENTA home page. Paid advertisements using the study video resulted in a total reach of 54,992 and 28,586 instances of page engagement. Post promotions, or paid advertisements that focused on increasing the reach of a post, resulted in 2666 reach and 97 instances of engagement. Conversion campaigns resulted in 268,052 reaches, 35,904 instances of engagement, and 9095 conversions. Campaigns seeking to drive traffic to the MAGENTA study website resulted in 80,120 reaches, 18,158 instances of engagement, and 1697 times that a unique user clicked on the link to the MAGENTA home page.

Almost all users engaged with paid advertisements from a handheld mobile device, such as a smartphone or tablet, rather than through a desktop computer. Most users engaged with paid advertisements from their Android device (35,806/373,682, 9.58%), followed by iOS devices (16,466/373,682, 4.41%) (eg, iPad). Approximately 9.98% (26,752/268,168) of the women aged <54 years reached by the advertisement content engaged with the advertisement, whereas approximately 26.62% (12,226/45,930) of women aged between 55 and 64 years who saw the paid content engaged, and 43.95% (15,124/34,410) of women aged ≥65 years who saw MAGENTA advertisements engaged with advertisement content. The difference observed between the above age demographics regarding reach to engagement was statistically significant ( $P<.001$ ). Unpaid posts published on the MAGENTA Facebook page resulted in 34,087 reaches and 3131 engagements. These numbers do not include social media posts published on other non-MAGENTA Facebook groups and pages.

**Table 3.** Global summary of results for all paid campaigns.

Advertisement name and objective	Reach	Post engagement, n (%)	Page engagement, n (%)	Video plays, n (%)	Link clicks, n (%)	Results, n (%)	Cost per results (US \$)
BCOC <sup>a</sup> Interest 8 and traffic	2280	43 (1.89)	46 (2.02)	N/A <sup>b</sup>	37 (1.62)	37 (1.62)	1.35
Latino 7 and traffic	11,062	217 (1.96)	223 (2.02)	N/A	204 (1.84)	204 (1.84)	1.23
BCOC Interest 6 and traffic	7536	1397 (18.54)	1397 (18.54)	N/A	266 (3.53)	266 (3.53)	0.94
BCOC Interest 5 and traffic	10,436	2069 (19.83)	2069 (19.83)	N/A	280 (2.68)	280 (2.68)	1.07
BCOC Interest 4 and traffic	34,648	6536 (18.86)	6536 (18.86)	N/A	640 (1.85)	540 (1.56)	1.94
BCOC Interest 9 and traffic	2050	351 (17.12)	351 (17.12)	N/A	20 (0.98)	20 (0.98)	2.00
African American and traffic	9382	7308 (77.89)	7308 (77.89)	N/A	319 (3.40)	319 (3.40)	0.63
Cascade Testing 1-2 and conversions	132,480	3341 (2.52)	3342 (2.52)	N/A	3326 (2.51)	2514 (1.90)	0.17
WA <sup>c</sup> and conversions	20,732	313 (1.51)	317 (1.53)	N/A	298 (1.44)	73 (0.35)	5.29
CA <sup>d</sup> 3 and conversions	35,976	2022 (5.62)	2022 (5.62)	N/A	2022 (5.62)	1692 (4.70)	0.21
NY <sup>e</sup> Campaign and traffic	395	186 (47.09)	186 (47.09)	N/A	14 (3.54)	14 (3.54)	0.53
Lookalike 1 and conversions	19,240	7142 (37.12)	7142 (37.12)	12,244 (63.64)	1130 (5.87)	880 (4.57)	0.40
BCOC Interest 1 and traffic	1587	31 (1.95)	33 (2.08)	N/A	7 (0.44)	7 (0.44)	3.77
BCOC Interest 2 and engagement	334	55 (16.47)	55 (16.47)	N/A	N/A	55 (16.47)	0.27
BCOC Interest 3 and traffic	746	10 (1.34)	10 (1.34)	N/A	10 (1.34)	10 (1.34)	1.50
CA 2 and conversions	35,752	21,444 (59.98)	21,444 (59.98)	7548 (21.11)	3147 (8.80)	2530 (7.08)	0.29
CA 1 and conversions	23,872	1637 (6.86)	1637 (6.86)	N/A	1637 (6.86)	1406 (5.89)	0.35

<sup>a</sup>BCOC: breast cancer and ovarian cancer.

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

<sup>c</sup>WA: Washington.

<sup>d</sup>CA: California.

<sup>e</sup>NY: New York.

## MAGENTA Study Enrollment and Randomization Summary

There were 34,715 unique visitors to the MD Anderson home page during the active social media recruitment period and 22,029 (63.46%) unique clicks. Approximately 63.46% (22,029/34,715) of users who visited the MD Anderson MAGENTA home page during this period clicked on the *Get Started* link, which directed them to the landing page on the REDCap system. The *Submit* button on the REDCap landing page received a total of 40.4% (14,025/34,715) of clicks, and the eligibility questionnaire on REDCap was completed 31.35% (10,883/34,715) of the times. Of the completed questionnaires, 14.02% (4887/34,715) were eligible. The enrollment and randomization data from the active social media recruitment period are summarized in [Table 2](#).

## Social Media Campaigns and News Stories Influence Enrollment Response

The general recruitment activity following paid advertisements was tracked and compared with periods when paid advertisements were not running. Because of the overlap in campaigns and television news stories, changes in recruitment activity around paid campaigns were not reported for all campaigns, and in some cases, the observation period following the campaign was excluded because of another campaign running during that time. There was an uptake in the completed eligibility questionnaires following individual and successive paid campaigns. Before 2 paid advertisements, which ran back to back in November 2017, there was a rate of 5.2 eligibility questionnaires completed daily. This number increased to a rate of 6.9 during these campaigns and in the week following the campaign. During the 2 weeks before another pair of paid advertisements, published sequentially in March 2018, there was a rate of 7.4 completed eligibility questionnaires daily, increasing to a rate of 12.7 during and in the 2 weeks following the campaign.

Enrollment following paid advertisement campaigns with a narrow geographical focus was further assessed. These campaigns included a targeted campaign in WA State (WA Campaign) and a campaign with multiple advertisements in California (CA; CA Campaign 1, CA Campaign 2, and CA Campaign 3), as seen in [Table 3](#). The WA Campaign reached 20,733 people, about 1.43% (298/20,733) of whom clicked on the webpage link, and 0.35% (73/20,733) went on to view content on the MD Anderson MAGENTA page. Throughout this campaign, a total of 32 individuals from WA State completed the eligibility questionnaire at a rate of 3.7 completed eligibility questionnaires per day. Before the social media campaign, a rate of 0.5 completed eligibility questionnaires daily.

The advertisement campaign targeting CA comprised 3 advertisements (CA Campaign 1, CA Campaign 2, and CA Campaign 3). This campaign had a reach of 95,600. Just >7% (6806/95,600, 7.12%) of these reaches resulted in a webpage link click, and some (5628/95,600, 5.89%) went on to view content on the MD Anderson MAGENTA page. During and after the immediate campaign, a total of 74 individuals from CA completed the eligibility questionnaire at a rate of 1.5

completed eligibility questionnaires per day. Before this campaign, there was a rate of 0.6 completed eligibility questionnaires per day from the state of CA.

During the active social media enrollment period, several television news stories, spearheaded by patient advocates and clinicians affiliated with the study, about the MAGENTA study were broadcast, including a story from WCCO based in Minnesota, [44], a Fox 2 Detroit story from Michigan, [45], and the King 5 story based in the WA State [46]. These news stories were widely shared over social media. In the month following the WCCO story, completed eligibility questionnaires from Minnesota increased from <0.5 per day to almost 123 per day. An increase in completed questionnaires was also observed following the release of the Fox 2 Detroit story. In the month immediately following this story, the number of completed eligibility questionnaires increased from 0.3 per day to 31 per day. Similarly, in the month following the King 5 story, completed eligibility questionnaires from WA State increased from 0.6 per day to 25 per day. These increases in enrollment and recruitment activities are shown in [Figure 5](#). Other increases, specifically those observed in study video views, aligned with paid Facebook advertising campaigns, where *video views* was the campaign objective.

## Discussion

### Principal Findings

This study demonstrated that Facebook is a useful way of reaching women aged >30 years who have a potentially increased risk of ovarian cancer through paid advertising, unpaid social media posts, and promoting news stories on social media. The key learning points include the following:

1. Campaign objectives that require more participant action to reach the end result generate passive engagement along the way.
2. Multimedia posts, specifically those with a video, create opportunities for engagement.
3. Effective social media outreach requires close collaboration with patient advocacy groups.
4. Web-based behavioral advertising can support targeted message delivery but is limited to those present on a specific platform.

In addition to these lessons, this research highlights other important limitations of social media outreach. Each of these learning points is addressed in greater detail below in the following sections.

### Campaign Objectives That Require More Action Generate Passive Engagement

More than one-quarter of the participants filling out the eligibility survey had heard about the study through social media, and another 28.16% (3938/13,983) through traditional media sources (ie, television news) that were then amplified by social media. Targeted, regional, paid Facebook advertising resulted in measurable increases in relevant regional enrollment for approximately 2 weeks following each campaign. These

recruitment sources were essential to the successful completion of MAGENTA enrollment and resulted in a wide national representation, with participants enrolling from all 50 states.

The engagement indicators reported across paid advertising were varied by the campaign (Table 3). The campaign objective, budget, schedule, duration, and targeted population all influenced the response rate and participant engagement. During the reported recruitment period, demographic targeting was modified by age, geographic location, and expressed interests on an ongoing basis. Campaigns that were more finely targeted by geographic location and prior engagement with cancer information or groups tended to cost more per result when compared with campaigns with broader targeting, presumably as the more customized population was comparatively smaller and more difficult to reach. Similarly, when the objective of the campaign required more action on the part of the participant to meet the objective, the cost per result increased. In other words, if the objective of the campaign was to get the participant to view the material on the study website, which would require the advertisement to appear on their screen, the participant must actually see the advertisement, click on the advertisement, go to the study home page, and spend a few moments with the study home page open on their browser. As a result, for example, this specific objective required a higher amount of engagement than a *post view* would. This also means that any advertisement with a multistep objective requiring more engagement accrued more upstream engagement. In the case of website views, to get a certain number of people to view the website, the Facebook advertising system required more people to see the initial post, spend time viewing that post, click the link, and so on. With this pattern in mind, we found that it was possible to increase post engagement upstream by focusing on downstream objectives that require more interaction to achieve. This incidental engagement also facilitated opportunities for repeated exposure, making it more likely for individual users to see information about the study more than once, potentially building brand recognition and familiarity.

### Multimedia Posts Create Opportunities for Engagement

Multimedia elements, such as the study video, were important for outreach during the study enrollment period. For example, Figure 5 depicts different ways that potential participants could interact with the web-based communication system, illustrating engagement with the study video, among other variables. The study video views fluctuated with the paid campaigns. Although many of the engagement increases observed in Figure 5 were connected to news stories and the subsequent boosting of these stories over social media, there were increases in study video views related to paid campaigns that had a video view objective. As we did not have a mechanism built into the web-based communication system that allowed us to determine how many participants learned about the study from watching the study video specifically, we were unable to calculate how many video views resulted in enrollment. Despite this limitation, video views likely helped build familiarity among the potential participants.

### Social Media Outreach Is Only as Strong as Your Relationships With Patient Advocates

This study also demonstrates the importance of including patient advocates as members of a multidisciplinary research team and using social media to boost patient advocate-spearheaded recruitment efforts. Patient advocacy groups supporting the MAGENTA study were critical to the success of the study. They not only helped facilitate televised and print news stories but also disseminated study information across their established web-based, as well as in-person, social networks. Importantly, patient advocates working with the study team also helped shape targeted advertising campaigns through Facebook's campaign targeting tools, which helped identify and boost content for individuals who followed patient advocacy Facebook pages.

Patient advocates were instrumental in designing accessible recruitment materials, getting news stories published, and supporting story circulation. Following the release of news stories featuring the MAGENTA study, there was a consistent increase in enrollment trends, with 28.16% (3938/13,983) of potential participants reporting that they learned about the study over television, referencing specific news stations featuring news clips about the study. These news segments, spearheaded by patient advocates, played a central role in the study recruitment. Although these stories originated via traditional media, either as televised news stories or similar publications, social media still likely played a role in promoting this content. Over social media, more people were able to view and share the news stories, making these news features more accessible. The inclusion of multimedia content, such as videos, appeared to extend this reach further, making web-based content easier to view and share. The advantage of video media is well-documented, with other research confirming that videos and other media-rich posts perform better than text-based content alone [47,48].

Given the spikes in page views and engagement that followed each news story, news stories were arguably one of the most effective outreach mechanisms used during the observed recruitment period. They are also one of the most difficult outreach mechanisms to implement, depending on either significant financial resources or existing interpersonal relationships with a news station or anchor. The MAGENTA study benefited from existing relationships between our patient advocate partners and local news anchors. If traditional media outreach such as this can be obtained, it can clearly be instrumental in meeting recruitment goals; however, it is unrealistic to count on it as a primary outreach mechanism. In addition, outreach that is geographically focused, such as news stories released over a specific network, will ultimately be limited to the demographic served by that network. This was certainly the case for the WCCO story, which is discussed in further detail in the following sections.

The WCCO televised story was arguably the most successful individual recruitment effort [44]. This story featured a local news anchor with a family history of breast and ovarian cancer named Kim Johnson. Johnson is an established household name for many of the communities served by the WCCO and has spoken publicly about ovarian cancer in the past. It is possible

that this story gained the traction it did for the same reasons that web-based information seekers are more likely to use familiar sources—if they can recognize the name, they are more likely to trust it [49]. Comparing these efforts with the enrollment activity following paid advertisements, it appears that although paid advertisements have an impact, collaboration with patient advocacy groups is also important for reaching a target audience. By leveraging existing social networks over social media through patient advocacy groups, Facebook could offer more cost-saving opportunities for research recruitment, particularly for large-scale studies such as MAGENTA. Considering these opportunities, as the average Facebook user continues to age [50], Facebook is likely to become an increasingly favorable venue for recruiting adults for research. A similar evolution in the average user is also observable across other social media platforms.

### Web-Based Behavioral Advertising Supports Message Delivery—But Not to Everyone

OBA made it easy to target information about the study to specific age groups, regions, and expressed and inferred interests. For example, we were able to target people who met the age and regional criteria and who had expressed an interest in various ovarian and breast cancer–related initiatives. Women who were aged  $\geq 65$  years had the best response rate when compared with other age groups, with approximately 43.95% (15,124/34,410) of reaches translating to engagement. This response rate suggests that although individuals aged  $\geq 65$  years make up a smaller percentage of web-based social media users, they are arguably more responsive to the content they see on social media than younger demographics. Their response rates could potentially be leveraged with a different message. Rather than encouraging them to enroll themselves, future advertisements might implore them to encourage their family members to learn more about the MAGENTA study.

Facebook and other social media platforms certainly present several opportunities for researchers; however, privacy concerns and worries over the use of OBA data make it clear that the drivers of Facebook do not always share the same values as the drivers of research. Paid advertising presents unique opportunities to target specific groups of people; however, unpaid posts published across existing web-based social networks are arguably preferable from an ethical standpoint, particularly with regard to recent data breaches on Facebook and concerns about how social media platforms such as Facebook use and monetize OBA data [51]. Data privacy issues such as these affect consumer trust and may deter users from previously trusted social media platforms, such as Facebook. Importantly, when unpaid posts come from existing social media profiles, such as a patient advocacy Facebook page with an established following, it is likely to function better than a sponsored advertisement, in large part because of this trust factor. When a message comes from a trusted source, patients are more likely to feel comfortable engaging with it. This requires research teams to build relationships with patient advocacy groups, specifically with those that include a following that meets the intended study eligibility criteria. In the absence of this invaluable resource, paid advertising may offer an effective alternative.

Most MAGENTA participants were White-identifying individuals. This may have been partly because of the geographic locations that recruitment bursts originated from; for example, the Minnesota burst increased enrollment from a region comprising  $>80\%$  non-Hispanic White individuals. Black and indigenous people of color are chronically underrepresented in clinical research settings [52]. This trend is partly explained by ineffective recruitment mechanisms [53]. The relatively homogenous sample recruited by the MAGENTA study poses a deficit for research, leaving underrepresented communities less likely to clinically benefit from research findings [52]. This problem is not unique to MAGENTA and is not something that social media recruitment alone can resolve.

Prior work suggests that different groups have different response rates where research is concerned, meaning that targeted marketing, even over social media, is likely to leave certain groups underrepresented [34]. Current recommendations highlight the importance of allowing the target population to inform platform choices [26]. Other social media platforms with sufficient Black and indigenous people of color representation should be explored for recruitment opportunities. Future research should assess the effectiveness of targeted recruitment across varying social media platforms for the purposes of reaching underrepresented populations and exploring alternative delivery models to improve access to genetic testing for Black and indigenous people of color communities.

A drop-off was observed from the initial engagement to enrollment and randomization (Table 2). The drop-off may be explained by a normal study drop-off at each stage; however, it may also be the result of the complex web-based enrollment protocol used. Participants who learned about the study were referred to the study webpage, and there, they were several clicks away from the eligibility questionnaire (Figure 1). Eligible individuals then had to note the messaging at the end of the questionnaire that told them to check their email inbox for an email containing the next steps and ensure that any auto-filtering system they had turned on in their email inbox did not filter the REDCap email directly into their trash or spam box. This issue came up during initial system usability testing and was addressed by adding additional messages at the end of the questionnaire, which prompted people to check their email inboxes. It is possible that some of the drop-off rates between the completion of the questionnaire and providing signed consent may be because of lost emails.

### Limitations

There were several limitations to this study. One of the most prominent issues was the varying definitions of *reach* and *engagement* across different web-based platforms. Although Facebook differentiated these variables, REDCap did not, making it difficult to accurately compare numbers across the various platforms included in the web-based communication system. This also made it difficult to determine whether a particular effort was successful or whether the participant finally took action after seeing information about the study for the third or fourth time, a potential trend supported by marketing research that indicates that repeated exposure is required for action [54]. Similarly, Facebook does not currently have a way of tracking



website conversions through unpaid posts or a public-facing means of tracking the demographics of engaged users or platforms from which they access content; thus, this information was not collected for unpaid published materials.

The platform itself also has a limitation. First, Facebook, similar to other social media platforms, is a rapidly evolving tool that uses internal user analytics to make changes to its terms and use agreements. This includes routine revisions of advertising platforms. For MAGENTA, this meant that some of the initial targeting variables and content used toward the beginning of the observed recruitment period were no longer available as the outreach continued. Although this is an issue that all social media platforms are likely to face, there are other reasons that researchers should carefully consider their options in social media platforms when choosing one for recruitment outreach.

Platform selection appears to be one of the most important factors in conducting social media research. The most popular social media platforms currently used for research recruitment are Facebook, Twitter, and Instagram. Each of these platforms has different user demographic profiles, with social media preferences varying by race, ethnicity, and age. Facebook is

increasingly becoming a platform that is appropriate for reaching middle-aged and older adults as the average Facebook user ages [18]. Instagram and Twitter, on the other hand, may be better options for reaching younger populations, given that the average Instagram and Twitter user is aged <35 years [18]. The average Twitter user, for example, is a young, affluent, college-educated male of color [18]. Certain racial-ethnic groups also tend toward other preferred social media platforms. For example, the most popular social media platform among Koreans is called KakaoTalk [55]. It is important to choose a social media platform populated by members of the intended population. This requires an understanding of the social media habits of the target population. It is also critical to understand that any social media platform will be subject to sampling bias if used to recruit research participants. Not only will recruitment activities be subject to the bias present on the specific platform but also be subject to the bias that results from internet-based recruitment efforts; that is, the resulting study population will largely be made up of individuals who use the internet, a potential marker of eHealth literacy and technology literacy. Regardless of the research goals, the target population should inform the social media platform choice.

---

## Acknowledgments

This work was supported by a Stand Up To Cancer (SU2C)-Ovarian Cancer Research Alliance-National Ovarian Cancer Coalition Ovarian Cancer Dream Team Translational Research grant (grant number SU2C-AACR-DT16-15). SU2C is a division of the Entertainment Industry Foundation. Research grants were administered by the American Association for Cancer Research, a scientific partner of SU2C.

---

## Conflicts of Interest

None declared.

---

## References

1. Reuter K, Angyan P, Le N, MacLennan A, Cole S, Bluthenthal RN, et al. Monitoring twitter conversations for targeted recruitment in cancer trials in Los Angeles county: protocol for a mixed-methods pilot study. *JMIR Res Protoc* 2018 Sep 25;7(9):e177 [FREE Full text] [doi: [10.2196/resprot.9762](https://doi.org/10.2196/resprot.9762)] [Medline: [30274964](https://pubmed.ncbi.nlm.nih.gov/30274964/)]
2. Strategies to improve recruitment to randomised trials. *Cochrane Library*. 2018. URL: <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.MR000013.pub6/full> [accessed 2022-04-10]
3. Briel M, Olu KK, von Elm E, Kasenda B, Alturki R, Agarwal A, et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *J Clin Epidemiol* 2016 Dec;80:8-15. [doi: [10.1016/j.jclinepi.2016.07.016](https://doi.org/10.1016/j.jclinepi.2016.07.016)] [Medline: [27498376](https://pubmed.ncbi.nlm.nih.gov/27498376/)]
4. Dew A, Khan S, Babinski C, Michel N, Heffernan M, Stephan S, et al. Recruitment strategy cost and impact on minority accrual to a breast cancer prevention trial. *Clin Trials* 2013 Apr;10(2):292-299 [FREE Full text] [doi: [10.1177/1740774512471452](https://doi.org/10.1177/1740774512471452)] [Medline: [23321266](https://pubmed.ncbi.nlm.nih.gov/23321266/)]
5. McIntosh S, Ossip-Klein DJ, Spada J, Burton K. Recruitment strategies and success in a multi-county smoking cessation study. *Nicotine Tob Res* 2000 Aug 1;2(3):281-284. [doi: [10.1080/14622200050147556](https://doi.org/10.1080/14622200050147556)] [Medline: [11082829](https://pubmed.ncbi.nlm.nih.gov/11082829/)]
6. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. *Annu Rev Public Health* 2006 Apr 01;27(1):1-28. [doi: [10.1146/annurev.publhealth.27.021405.102113](https://doi.org/10.1146/annurev.publhealth.27.021405.102113)] [Medline: [16533107](https://pubmed.ncbi.nlm.nih.gov/16533107/)]
7. Caplan A, Friesen P. Health disparities and clinical trial recruitment: is there a duty to tweet? *PLoS Biol* 2017 Mar 1;15(3):e2002040 [FREE Full text] [doi: [10.1371/journal.pbio.2002040](https://doi.org/10.1371/journal.pbio.2002040)] [Medline: [28249024](https://pubmed.ncbi.nlm.nih.gov/28249024/)]
8. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. *J Med Internet Res* 2016 Nov 07;18(11):e286 [FREE Full text] [doi: [10.2196/jmir.5698](https://doi.org/10.2196/jmir.5698)] [Medline: [27821383](https://pubmed.ncbi.nlm.nih.gov/27821383/)]
9. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. *J Med Internet Res* 2012 Feb 01;14(1):e20 [FREE Full text] [doi: [10.2196/jmir.1978](https://doi.org/10.2196/jmir.1978)] [Medline: [22297093](https://pubmed.ncbi.nlm.nih.gov/22297093/)]

10. Temple E, Brown RF. A comparison of internet-based participant recruitment methods: engaging the hidden population of cannabis users in research. *J Res Pract* 2011;7(2) [FREE Full text]
11. Obar J, Wildman SS. Social media definition and the governance challenge: an introduction to the special issue. *Telecommun Policy* 2015;39(9):745-750. [doi: [10.2139/ssrn.2637879](https://doi.org/10.2139/ssrn.2637879)]
12. Dizon DS, Graham D, Thompson MA, Johnson LJ, Johnston C, Fisch MJ, et al. Practical guidance: the use of social media in oncology practice. *J Oncol Pract* 2012 Sep;8(5):e114-e124. [doi: [10.1200/jop.2012.000610](https://doi.org/10.1200/jop.2012.000610)]
13. Gorman K, Smith AM, Cimini ME, Halloran KM, Lubiner AG. Reaching the hard to reach: lessons learned from a statewide outreach initiative. *J Commun Practice* 2013 Jan;21(1-2):105-123 [FREE Full text] [doi: [10.1080/10705422.2013.788369](https://doi.org/10.1080/10705422.2013.788369)]
14. Seltzer E, Stolley MR, Mensah EK, Sharp LK. Social networking site usage among childhood cancer survivors--a potential tool for research recruitment? *J Cancer Surviv* 2014 Sep;8(3):349-354 [FREE Full text] [doi: [10.1007/s11764-014-0348-4](https://doi.org/10.1007/s11764-014-0348-4)] [Medline: [24532046](https://pubmed.ncbi.nlm.nih.gov/24532046/)]
15. Yuan P, Bare MG, Johnson MO, Saberi P. Using online social media for recruitment of human immunodeficiency virus-positive participants: a cross-sectional survey. *J Med Internet Res* 2014 May 01;16(5):e117 [FREE Full text] [doi: [10.2196/jmir.3229](https://doi.org/10.2196/jmir.3229)] [Medline: [24784982](https://pubmed.ncbi.nlm.nih.gov/24784982/)]
16. Rabin C, Horowitz S, Marcus B. Recruiting young adult cancer survivors for behavioral research. *J Clin Psychol Med Settings* 2013 Mar;20(1):33-36 [FREE Full text] [doi: [10.1007/s10880-012-9317-0](https://doi.org/10.1007/s10880-012-9317-0)] [Medline: [22810954](https://pubmed.ncbi.nlm.nih.gov/22810954/)]
17. Number of Facebook users in the United States from 2017 to 2026. Statista. URL: <https://www.statista.com/statistics/408971/number-of-us-facebook-users/> [accessed 2022-04-10]
18. Social media use in 2021. Pew Research Center. URL: <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/> [accessed 2022-04-10]
19. Chiang A, Vartabedian B, Spiegel B. Harnessing the hashtag: a standard approach to GI dialogue on social media. *Am J Gastroenterol* 2016 Aug;111(8):1082-1084. [doi: [10.1038/ajg.2016.259](https://doi.org/10.1038/ajg.2016.259)] [Medline: [27349338](https://pubmed.ncbi.nlm.nih.gov/27349338/)]
20. Curtis B. Social networking and online recruiting for HIV research: ethical challenges. *J Empir Res Hum Res Ethics* 2014 Feb;9(1):58-70 [FREE Full text] [doi: [10.1525/jer.2014.9.1.58](https://doi.org/10.1525/jer.2014.9.1.58)] [Medline: [24572084](https://pubmed.ncbi.nlm.nih.gov/24572084/)]
21. Curtis B. Understanding tailored internet smoking cessation messages. *ANNALS Am Acad Polit Soc Sci* 2012 Feb 10;640(1):136-149 [FREE Full text] [doi: [10.1177/0002716211423642](https://doi.org/10.1177/0002716211423642)]
22. Voytek C, Jones K, Curtis B, Fiore D, Dunbar D, Frank I, et al. Using an internet consumer marketing strategy to reach men who have sex with men for participation in a preventive HIV vaccine clinical trial. *Retrovirology* 2012 Sep 13;9(S2):119. [doi: [10.1186/1742-4690-9-s2-p119](https://doi.org/10.1186/1742-4690-9-s2-p119)]
23. Bull S, Breslin LT, Wright EE, Black SR, Levine D, Santelli JS. Case study: an ethics case study of HIV prevention research on Facebook: the just/us study. *J Pediatr Psychol* 2011;36(10):1082-1092 [FREE Full text] [doi: [10.1093/jpepsy/jsq126](https://doi.org/10.1093/jpepsy/jsq126)] [Medline: [21292724](https://pubmed.ncbi.nlm.nih.gov/21292724/)]
24. Carpenter M, Sterba KR, Boatright AS, West R. 'Closet' quit attempts: prevalence, correlates and association with outcome. *Addiction* 2011 Dec;106(12):2214-2220 [FREE Full text] [doi: [10.1111/j.1360-0443.2011.03538.x](https://doi.org/10.1111/j.1360-0443.2011.03538.x)] [Medline: [21672072](https://pubmed.ncbi.nlm.nih.gov/21672072/)]
25. Hagan A. The influence of support groups and identity on goal setting and achievement in persons with type 2 diabetes. University of Michigan Library. 2010. URL: <https://deepblue.lib.umich.edu/handle/2027.42/75845> [accessed 2022-04-10]
26. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. *Digit Health* 2018 May 07;4:2055207618771757 [FREE Full text] [doi: [10.1177/2055207618771757](https://doi.org/10.1177/2055207618771757)] [Medline: [29942634](https://pubmed.ncbi.nlm.nih.gov/29942634/)]
27. Mostaghimi A, Crotty BH. Professionalism in the digital age. *Ann Intern Med* 2011 Apr 19;154(8):560-562 [FREE Full text] [doi: [10.7326/0003-4819-154-8-201104190-00008](https://doi.org/10.7326/0003-4819-154-8-201104190-00008)] [Medline: [21502653](https://pubmed.ncbi.nlm.nih.gov/21502653/)]
28. Greysen S, Kind T, Chretien KC. Online professionalism and the mirror of social media. *J Gen Intern Med* 2010 Nov;25(11):1227-1229 [FREE Full text] [doi: [10.1007/s11606-010-1447-1](https://doi.org/10.1007/s11606-010-1447-1)] [Medline: [20632121](https://pubmed.ncbi.nlm.nih.gov/20632121/)]
29. Baltar F, Brunet I. Social research 2.0: virtual snowball sampling method using Facebook. *Internet Res* 2012 Jan 27;22(1):57-74. [doi: [10.1108/10662241211199960](https://doi.org/10.1108/10662241211199960)]
30. Brickman Bhutta C. Not by the book. *Sociol Method Res* 2012 Mar 21;41(1):57-88 [FREE Full text] [doi: [10.1177/0049124112440795](https://doi.org/10.1177/0049124112440795)]
31. Using social media for clinical trial recruitment. Pharmaphorum. URL: <https://pharmaphorum.com/views-and-analysis/using-social-media-for-clinical-trial-recruitment/> [accessed 2022-04-10]
32. Dosek T. Snowball sampling and Facebook: how social media can help access hard-to-reach populations. *PS Political Sci Politics* 2021 May 18;54(4):651-655. [doi: [10.1017/S104909652100041X](https://doi.org/10.1017/S104909652100041X)]
33. Griffiths F, Cave J, Boardman F, Ren J, Pawlikowska T, Ball R, et al. Social networks--the future for health care delivery. *Soc Sci Med* 2012 Dec;75(12):2233-2241 [FREE Full text] [doi: [10.1016/j.socscimed.2012.08.023](https://doi.org/10.1016/j.socscimed.2012.08.023)] [Medline: [22985490](https://pubmed.ncbi.nlm.nih.gov/22985490/)]
34. 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](https://doi.org/10.2196/jmir.7071)] [Medline: [28851679](https://pubmed.ncbi.nlm.nih.gov/28851679/)]
35. Kapp J, Peters C, Oliver DP. Research recruitment using Facebook advertising: big potential, big challenges. *J Cancer Educ* 2013 Mar;28(1):134-137 [FREE Full text] [doi: [10.1007/s13187-012-0443-z](https://doi.org/10.1007/s13187-012-0443-z)] [Medline: [23292877](https://pubmed.ncbi.nlm.nih.gov/23292877/)]

36. Bajardi P, Paolotti D, Vespignani A, Eames K, Funk S, Edmunds WJ, et al. Association between recruitment methods and attrition in internet-based studies. *PLoS One* 2014;9(12):e114925 [FREE Full text] [doi: [10.1371/journal.pone.0114925](https://doi.org/10.1371/journal.pone.0114925)] [Medline: [25490045](https://pubmed.ncbi.nlm.nih.gov/25490045/)]
37. Frandsen M, Walters J, Ferguson SG. Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. *Nicotine Tob Res* 2014 Feb;16(2):247-251 [FREE Full text] [doi: [10.1093/ntr/ntt157](https://doi.org/10.1093/ntr/ntt157)] [Medline: [24127266](https://pubmed.ncbi.nlm.nih.gov/24127266/)]
38. Swisher EM, Rayes N, Bowen D, Peterson CB, Coffin T, Norquist B, et al. Results from MAGENTA: a national randomized four-arm noninferiority trial evaluating pre- and post-test genetic counseling during online testing for breast and ovarian cancer genetic risk. *J Clin Oncol* 2020 May 20;38(15\_suppl):1506. [doi: [10.1200/jco.2020.38.15\\_suppl.1506](https://doi.org/10.1200/jco.2020.38.15_suppl.1506)]
39. 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](https://doi.org/10.2196/jmir.5502)] [Medline: [27306780](https://pubmed.ncbi.nlm.nih.gov/27306780/)]
40. Musiat P, Winsall M, Orlowski S, Antezana G, Schrader G, Battersby M, et al. Paid and unpaid online recruitment for health interventions in young adults. *J Adolesc Health* 2016 Dec;59(6):662-667. [doi: [10.1016/j.jadohealth.2016.07.020](https://doi.org/10.1016/j.jadohealth.2016.07.020)] [Medline: [27663927](https://pubmed.ncbi.nlm.nih.gov/27663927/)]
41. O'Connor A, Jackson L, Goldsmith L, Skirton H. Can I get a retweet please? Health research recruitment and the Twittersphere. *J Adv Nurs* 2014 Mar 04;70(3):599-609. [doi: [10.1111/jan.12222](https://doi.org/10.1111/jan.12222)] [Medline: [23909740](https://pubmed.ncbi.nlm.nih.gov/23909740/)]
42. Walker D. The internet as a medium for health services research Part 2. *Nurse Res* 2013 May;20(5):33-37. [doi: [10.7748/nr2013.05.20.5.33.e295](https://doi.org/10.7748/nr2013.05.20.5.33.e295)] [Medline: [23687847](https://pubmed.ncbi.nlm.nih.gov/23687847/)]
43. Batterham P. Recruitment of mental health survey participants using internet advertising: content, characteristics and cost effectiveness. *Int J Methods Psychiatr Res* 2014 Jun;23(2):184-191 [FREE Full text] [doi: [10.1002/mpr.1421](https://doi.org/10.1002/mpr.1421)] [Medline: [24615785](https://pubmed.ncbi.nlm.nih.gov/24615785/)]
44. Trial offers at-home test kit to show risk of ovarian cancer. CBS Minnesota. URL: <https://minnesota.cbslocal.com/2018/04/30/ovarian-cancer-magenta-test-national-trial/> [accessed 2022-04-10]
45. MAGENTA free genetic testing helps woman avoid cancer. FOX 2 Detroit. URL: <https://www.fox2detroit.com/news/magenta-free-genetic-testing-helps-woman-avoid-cancer> [accessed 2022-04-10]
46. Magenta genetic test can screen for ovarian cancer. KING 5. URL: <https://www.king5.com/article/news/health/magenta-genetic-test-can-screen-for-ovarian-cancer/281-586362861> [accessed 2022-04-10]
47. Ordenes F, Grewal D, Ludwig S, Ruyter KD, Mahr D, Wetzels M. Cutting through content clutter: how speech and image acts drive consumer sharing of social media brand messages. *J Consum Res* 2019;45(5):988-1012. [doi: [10.1093/jcr/ucy032](https://doi.org/10.1093/jcr/ucy032)]
48. Theiss SK, Burke RM, Cory JL, Fairley TL. Getting beyond impressions: an evaluation of engagement with breast cancer-related Facebook content. *Mhealth* 2016 Nov 07;2:41 [FREE Full text] [doi: [10.21037/mhealth.2016.10.02](https://doi.org/10.21037/mhealth.2016.10.02)] [Medline: [27840816](https://pubmed.ncbi.nlm.nih.gov/27840816/)]
49. Steelman TA, McCaffrey SM, Velez AK, Briefel JA. What information do people use, trust, and find useful during a disaster? Evidence from five large wildfires. *Nat Hazards* 2014 Nov 16;76(1):615-634. [doi: [10.1007/s11069-014-1512-x](https://doi.org/10.1007/s11069-014-1512-x)]
50. Allen K. The average Facebook user is getting older—and more masculine. PR Daily. URL: <https://www.prdaily.com/the-average-facebook-user-is-getting-older-and-more-masculine/> [accessed 2022-04-10]
51. Mark Zuckerberg Testimony: senators question Facebook's commitment to privacy. The New York Times. URL: <https://www.nytimes.com/2018/04/10/us/politics/mark-zuckerberg-testimony.html> [accessed 2022-04-10]
52. Davis TC, Arnold CL, Mills G, Miele L. A qualitative study exploring barriers and facilitators of enrolling underrepresented populations in clinical trials and biobanking. *Front Cell Dev Biol* 2019 Apr 30;7:74 [FREE Full text] [doi: [10.3389/fcell.2019.00074](https://doi.org/10.3389/fcell.2019.00074)] [Medline: [31114788](https://pubmed.ncbi.nlm.nih.gov/31114788/)]
53. Russomanno J, Patterson JG, Jabson Tree JM. Social media recruitment of marginalized, hard-to-reach populations: development of recruitment and monitoring guidelines. *JMIR Public Health Surveill* 2019 Dec 02;5(4):e14886 [FREE Full text] [doi: [10.2196/14886](https://doi.org/10.2196/14886)] [Medline: [31789598](https://pubmed.ncbi.nlm.nih.gov/31789598/)]
54. Tripp C, Jensen TD, Carlson L. The effects of multiple product endorsements by celebrities on consumers' attitudes and intentions. *J Consum Res* 1994 Mar;20(4):535. [doi: [10.1086/209368](https://doi.org/10.1086/209368)]
55. Jin DY, Yoon K. Reimagining smartphones in a local mediascape. *Convergence* 2016 Jul 07;22(5):510-523. [doi: [10.1177/1354856514560316](https://doi.org/10.1177/1354856514560316)]

## Abbreviations

- CA:** California
- MAGENTA:** Making Genetic Testing Accessible
- OBA:** online behavioral advertising
- REDCap:** Research Electronic Data Capture
- SU2C:** Stand Up To Cancer
- WA:** Washington

*Edited by A Mavragani; submitted 17.11.21; peer-reviewed by H Kim, D Elling, J Zhao; comments to author 24.01.22; revised version received 08.03.22; accepted 16.03.22; published 26.09.22.*

*Please cite as:*

*Coffin T, Bowen D, Lu K, Swisher EM, Rayes N, Norquist B, Blank SV, Levine DA, Bakkum-Gamez JN, Fleming GF, I Olopade O, Romero I, D'Andrea A, Nebgen DR, Peterson C, Munsell MF, Gavin K, Crase J, Polinsky D, Lechner R*

*Using Social Media to Facilitate Communication About Women's Testing: Tool Validation Study*

*JMIR Form Res 2022;6(9):e35035*

*URL: <https://formative.jmir.org/2022/9/e35035>*

*doi: [10.2196/35035](https://doi.org/10.2196/35035)*

*PMID: [36155347](https://pubmed.ncbi.nlm.nih.gov/36155347/)*

©Tara Coffin, Deborah Bowen, Karen Lu, Elizabeth M Swisher, Nadine Rayes, Barbara Norquist, Stephanie V Blank, Douglas A Levine, Jamie Nadine Bakkum-Gamez, Gini F Fleming, Olufunmilayo I Olopade, Iris Romero, Alan D'Andrea, Denise R Nebgen, Christine Peterson, Mark F Munsell, Kathleen Gavin, Jamie Crase, Deborah Polinsky, Rebecca Lechner. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 26.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# An Economic Impact Model for Estimating the Value to Health Systems of a Digital Intervention for Diabetes Primary Care: Development and Usefulness Study

---

Brenton Powers<sup>1</sup>, MBA; Amy Bucher<sup>1</sup>, PhD

Lirio, Knoxville, TN, United States

**Corresponding Author:**

Amy Bucher, PhD

Lirio

320 Corporate Drive, NW

Knoxville, TN, 37923

United States

Phone: 1 865 839 6158

Email: [abucher@lirio.com](mailto:abucher@lirio.com)

## Abstract

**Background:** Diabetes is associated with significant long-term costs for both patients and health systems. Regular primary care visits aligned with American Diabetes Association guidelines could help mitigate those costs while generating near-term revenue for health systems. Digital interventions prompting primary care visits among unengaged patients could provide significant economic value back to the health system as well as individual patients, but only few economic models have been put forth to understand this value.

**Objective:** Our objective is to establish a data-based method to estimate the economic impact to a health system of interventions promoting primary care visits for people with diabetes who have been historically unengaged with their care. The model was built with a focus on a specific digital health intervention, Precision Nudging, but can be used to quantify the value of other interventions driving primary care usage among patients with diabetes.

**Methods:** We developed an economic model to estimate the financial value of a primary care visit of a patient with diabetes to the health system. This model requires segmenting patients with diabetes according to their level of blood sugar control as measured by their most recent hemoglobin A<sub>1c</sub> value to understand how frequently they should be visiting a primary care provider. The model also accounts for the payer mix among the population with diabetes, documenting the percentage of insurance coverage through a commercial plan, Medicare, or Medicaid, as these influence the reimbursement rates for the services. Then, the model takes into consideration the population base rates of comorbid conditions for patients with diabetes and the associated current procedural terminology codes to understand what a provider can bill as well as the expected inpatient revenue from a subset of patients likely to require hospitalization based on the national hospitalization rates for people with diabetes. Physician reimbursement is subtracted from the total. Finally, the model also accounts for the level of patient engagement with the intervention to ensure a realistic estimate of the impact.

**Results:** We present a model to prospectively estimate the economic impact of a digital health intervention to encourage patients with documented diabetes diagnoses to attend primary care visits. The model leverages both publicly available and health system data to calculate the per appointment value (revenue) to the health system. The model offers a method to understand and test the financial impact of Precision Nudging or other primary care-focused diabetes interventions inclusive of costs driven by comorbid conditions.

**Conclusions:** The proposed economic model can help health systems understand and evaluate the estimated economic benefits of interventions focused on primary care and prevention for patients with diabetes as well as help intervention developers determine pricing for their product.

(*JMIR Form Res* 2022;6(9):e37745) doi:[10.2196/37745](https://doi.org/10.2196/37745)

## KEYWORDS

return on investment; value; payment model; cost; economic impact; digital health; eHealth; diabetes; primary care; email

## Introduction

The Centers for Disease Control and Prevention (CDC) estimates that in 2018, 34.2 million Americans or 10.5% of the US population had diabetes [1]. The economic costs associated with diabetes in the United States are estimated to be in excess of US \$327 billion total, with US \$237 billion coming from direct medical costs and the rest in productivity reductions. For individual patients, the economic impact of diabetes can be catastrophic, with estimated annual medical expenses of US \$9600 directly attributable to diabetes and a total of US \$16,750 on average, with medical expenditures exceeding 2.3 times the amount incurred by patients without diabetes [2]. People with diabetes are also likely to have comorbid conditions that contribute significantly to their medical expenses [3,4], with 40% of adults with diabetes having at least 3 comorbid chronic diseases [5]. Moreover, people with diabetes have an increased odds of inpatient admission [6], with 34% of patient admissions occurring among people with diabetes [1] and a higher likelihood of readmission within 30 days of discharge [7] compared to people without diabetes (20.5%) [8]. In short, diabetes is both prevalent and expensive.

It is also common for people with diabetes to not be aware that they have the condition, which limits their ability to engage in appropriate care. In 2018, 21.4% of US adults with diabetes did not report having the condition—a total of 7.3 million people whose laboratory results qualify for a diagnosis [1]. This group is almost certainly not engaged in recommended condition management behaviors such as primary care and specialist appointments specifically to address diabetes. Engagement is lacking in people who are aware of their diabetes status as well; for example, one study on people with diabetes found 16.2% were no-shows to their last scheduled primary care appointment [9]. In general, missed medical appointments are estimated to cost the US health care system in excess of US \$150 billion per year [10], indicating a need for interventions that increase patient attendance.

Noncompliance with recommended care has serious consequences for both individuals and systems. A comparison of people with diabetes who were compliant and noncompliant with American Diabetes Association recommendations for primary care found significant improvements in medical utilization, including a reduction in the need for inpatient care [11] when recommendations were followed. For example, one study found that promoting lifestyle changes around diet and physical activity for people with diabetes and prediabetes yielded significant cost savings related to medical care over a 10-year time period [12]. Regular well visits with providers are also associated with positive outcomes for people with diabetes. Research suggests that about 30% of patients working with a primary care provider (PCP) achieve hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>) control by 1 year [13], and in general, interventions targeting appointment attendance are associated with better diabetes outcomes [14].

Naturally, many behavioral interventions for people with diabetes focus on adherence to recommended clinical pathways to improve outcomes. In the digital health space, these

interventions often focus on blood glucose monitoring, diabetes education, and lifestyle modifications [15] as the key modifiable behaviors. In general, these digital health apps have heterogeneous outcomes with some promise to help lower HbA<sub>1c</sub> levels and other biometrics as well as improve patient self-efficacy and condition management skills [16]. For example, a meta-analysis and evidence review of app-based interventions for type 2 diabetes suggests an overall reduction in HbA<sub>1c</sub> levels among users, particularly among younger users, and when there is a feedback loop involving the provider [17]. A number of these digital health interventions claim to reduce the system- and individual-level costs of diabetes. For example, in one longitudinal claims-based study of Omada Diabetes Prevention Program, participants incurred reduced health care costs at 1 year after enrollment, including reduced inpatient costs [18]. Evidence suggests that digital health interventions may be an effective tool to augment clinical-based diabetes care, but there remain gaps in the evidence base, particularly around the economic impact [19,20] and a relative dearth of interventions focused on supporting traditional clinical pathways such as primary care. Finally, models focused on cost reduction may overlook the value to health systems from revenue generation via primary care utilization.

In short, diabetes is a complex condition that can drive costs from a number of sources, whether through more intensive preventative care needs, frequently comorbid conditions, or complications or sequelae requiring more expensive treatment. Given the focus of most diabetes digital health interventions, estimates of the value of these interventions often focus on the impact of lifestyle changes rather than appropriate utilization of preventative care such as regular PCP appointments. Primary prevention and regular care are likely to deliver value to the health system beyond their direct impact and have been considered as components of a value-based agreement for diabetes care [21,22]. We believe there is cause to focus on primary care and well visits as a modifiable behavior for people with diabetes and a need to quantify the economic value of doing so in order to appropriately prioritize interventions.

In this paper, we put forth a conceptual model and process to prospectively estimate the downstream economic benefits of a specific primary care-focused diabetes intervention, Precision Nudging for diabetes, to a health system. This model will ultimately provide the basis for assessing the intervention's value postimplementation, with a focus on revenue to the health system rather than a specific clinical outcome or quality-adjusted life years [23]. Although the model was developed to assess a specific intervention, we believe it is generalizable to understand the economic impact of other digital health initiatives focused on primary care utilization for people with diabetes. This would permit value-based pricing of commercial digital health interventions as well as an evidence-based method for health systems and provider organizations to determine whether and how to utilize such interventions for their populations.

## Methods

### Overview

The model was developed prospectively to quantify the impact of a specific diabetes behavioral intervention called as Precision Nudging. The key outcome measure for the intervention is the number of patients who attend a PCP appointment; therefore, the model focuses on assigning a financial value to each appointment to understand the financial impact of the intervention. The model focuses on the value that accrues to a health system, taking into account direct and indirect costs of diabetes as well as provider reimbursements, and was specifically intended to help health systems evaluate the value of Precision Nudging for their population.

### Precision Nudging Intervention

Precision Nudging for diabetes is an English-language messaging intervention for people with a diagnosis of type 1 or type 2 diabetes, focused on the target behaviors of scheduling and attending a well visit with a PCP either once, twice, or 4 times per year based on their most recent  $A_{1c}$  value as recommended by the Healthcare Effectiveness Data and Information Set (HEDIS) [24]. A behavioral reinforcement learning algorithm [25-27] selects behavioral science-based message content to send to eligible patients to prompt them to schedule and attend a diabetes well visit. The algorithm is designed to select messages based on recipient characteristics with an emphasis on prior recipient behavioral responses (ie, message opens and clicks and appointment scheduling and attendance), to maximize the likelihood the message is opened and the call to action heeded.

The messages are designed to address specific barriers people might have to scheduling and attending a diabetes well visit, identified through primary research during the intervention development process and a comprehensive literature review. Behavioral designers use an intervention mapping process to categorize the determinants and align them with behavior change techniques [28,29], which are then operationalized as message content (subject lines and body copy) and visual designs. The behavioral reinforcement learning algorithm compiles a complete message for each recipient from 10 subject lines and 26 body copy/visual design options, for a total of 320 unique message combinations that a patient might receive. Eligible patients receive 1 email per week for 5 weeks, followed by an 8-week pause, and then another 1 message per week for 5 weeks. This pattern repeats until the patient either unsubscribes from the intervention or takes action by scheduling and attending a PCP visit. The key outcome metric associated with Precision Nudging is the completion of a primary care appointment. The outcome of appointment attendance serves as the basis for the economic model.

### Establishment of Diabetes Impact to the Health System

A best practice in developing economic impact models is to clearly identify the entity to which the value is delivered [30]. In this case, it is the health system. The first step in the economic model is to establish the impact of diabetes to the health system with respect to diabetes severity and risk level, as characterized by patients' most recent  $HbA_{1c}$  value. This analysis will vary by health system depending on their patient population and payer mix. First, an understanding of the health system's patient population with diabetes must be established. This occurs by parsing electronic medical record data to identify patients with a documented diagnosis of diabetes (excluding gestational or medication-induced diabetes). The method of documenting a diabetes diagnosis may vary depending on system and implementation. In our case example, the health system maintained a diabetes registry within its Epic implementation that provided the base estimates for diabetes in the patient population.

From there, we segmented patients with diabetes based on their last recorded  $HbA_{1c}$  value. Following Healthcare Effectiveness Data and Information Set measures [24], patients were grouped as well-controlled if their  $HbA_{1c}$  level was below 7, moderately controlled if their  $HbA_{1c}$  level was between 7 and 9, and uncontrolled if their  $HbA_{1c}$  level was above 9. Patients who did not have an  $HbA_{1c}$  value in their records were classified as moderately well-controlled for messaging purposes, with the rationale that their PCP would order  $A_{1c}$  testing in a first or second visit and then the patient would be classified based on their actual value. In addition to facilitating guideline-consistent recommendations for care within the intervention, segmenting patients by  $HbA_{1c}$  value also permits the risk adjustment that has been identified as a best practice for understanding the value [31].

Because the intervention targets patients who are not up-to-date with recommended PCP appointments, we also set eligibility parameters based on the date of the last primary care visit. The recommended frequency of primary care visits differs by  $HbA_{1c}$  level; therefore, well-controlled patients are recommended 1 visit per year, moderately well-controlled 2 visits per year, and poorly controlled 4 visits per year, consistent with American Diabetes Association guidelines [32]. We classify patients who have not had a primary care visit in the appropriate lookback period and do not have one scheduled in the next 3 months as unengaged and consider them eligible for intervention messaging.

In working with a health system, it may be necessary to repeat the exercise of identifying patients with diabetes and classifying them by  $HbA_{1c}$  level per market or care site where the intervention will be offered, particularly if the payer mix or provider reimbursements vary by location. Table 1 offers an example of what this documentation may look like per location.

**Table 1.** Quantification of the population with diabetes within a health system or market by the hemoglobin A<sub>1c</sub> value and being unengaged with care due to not having a past appointment within the recommended time frame or a future visit scheduled.

Patient HbA <sub>1c</sub> <sup>a</sup> level	Criterion 1 (HbA <sub>1c</sub> value)	Patients (n)	And criterion 2 (appointment overdue)	Patients (n)	And criterion 3 (next appointment more than x time in future)	Total unengaged population
Controlled	HbA <sub>1c</sub> <7	XXX	Last appointment >11 months	XXX	3 months	XXX
Moderately controlled	HbA <sub>1c</sub> ≥7 to <9	XXX	Last appointment >5 months	XXX	3 months	XXX
Uncontrolled	HbA <sub>1c</sub> ≥9	XXX	Last appointment >2 months	XXX	3 months	XXX

<sup>a</sup>HbA<sub>1c</sub>: hemoglobin A<sub>1c</sub>.

### Identification of Current Procedural Terminology Codes

The next step in developing the model was to identify the current procedural terminology (CPT) codes corresponding to the procedures that may take place during a primary care visit for a patient with diabetes. Then, each code was assigned an allocation based on the percentage of patients who would be expected to receive that code on a given visit. For example, every PCP visit merits a CPT code for physician office visit; therefore, that code receives a 100% allocation, while the CPT code for a lipid panel is assigned a 44% allocation based on the national rate of hypercholesterolemia among people with diabetes [1]. A total of 10 CPT codes were identified and assigned an allocation percentage (see Table 2). The identified codes include professional services only that can be bundled

within the parameters of a physician visit. Hospital laboratory services related to the visit were excluded.

CPT code volumes were then adjusted based on the percentage of patients in the health system with a status of well controlled, moderately controlled, or uncontrolled and the corresponding number of annual wellness visits recommended (1, 2, or 4, respectively). Finally, the health system's patient population was characterized in terms of its payer mix to assign a financial value to each CPT code and its projected frequency. Dollar amounts were assigned based on typical reimbursement rates for each CPT code by plan, as determined by (1) Medicare: based on data from Palmetto GBA, (2) Medicaid: based on the rates for the state where the health system is located, and (3) commercial health plan: based on an average 135% of the Medicare reimbursement.

**Table 2.** The 10 current procedural terminology codes and their corresponding allocations based on the percentage of patients with diabetes likely to need them in a primary care provider appointment in order to determine the revenue potential of each appointment.

Current procedural terminology code description	Allocation (%)
Physician office visit	100
Hemoglobin A <sub>1c</sub> level	100
Urinalysis	100
Lipid panel	44
Complete blood count with auto-differential	100
Education on self-managed blood pressure setup	68
Education on self-managed blood pressure monitor	68
Tobacco cessation	22
Diabetic foot examination	100
Depression	25

### Physician Reimbursement

The revenue to the professional practice is offset by provider reimbursement. Physicians are reimbursed based on the CPT codes they submit, with each code having a relative value unit assigned to it by the Centers for Medicare and Medicaid Services (CMS). The worked relative value units are multiplied by a conversion factor, which may include a geographical adjustment, to arrive at a dollar value for each service. In order to account for physician reimbursement in the model, provider reimbursement was estimated using the median Medical Group

Management Association conversion factor on CMS relative value units for the CPT codes charged [33].

### Inpatient Care Reimbursement

People with diabetes are more likely to be admitted to inpatient care than people without diabetes [6]; the CDC reports that 339 out of 1000 people with diabetes may experience the need for inpatient care over a 1-year period [1]. Fortunately, regular PCP visits may reduce the risk of inpatient care, as following American Diabetes Association guidelines is associated with decreased admission rates [11]. Because the Precision Nudging



intervention focused on the unengaged population who were out of compliance with the recommended cadence of primary care visits, we focused on only that group in calculating potential inpatient care costs. Revenue is calculated on the margin from reimbursement by commercial health plans, Medicare, and Medicaid in the proportion those payers cover the unengaged patients exist in the system’s population. These calculations yield a value summary.

**Summary Calculation**

When applied to the population of patients with diabetes within a health system, this economic impact model yields an annual incremental value summary for the intervention—that is, a specific dollar value per primary care visit scheduled as a result of the intervention. The total margin is calculated by adding total professional practice revenue and total hospital margin and subtracting total provider compensation. This amount can be annualized and divided by the number of visits attended to arrive at a value per visit. This dollar amount supports the calculation of a return on investment based on the costs of implementation and operational support for the intervention.

The final calculation using the economic impact model with the inputs described above is as follows:

$$\text{Professional practice revenue (annual) + inpatient revenue (annual) - provider compensation (annual) = Total annual reimbursement to health system} / \text{Total attended PCP appointments from 100,000 patients} = \text{Dollar value per appointment}$$

**Ethical Considerations**

The development of the economic impact model did not utilize human subjects and so was not submitted for institutional review board review.

**Results**

Although we are unable to provide the specific calculation used with the health system implementation due to its use of proprietary information, the following example illustrates the process with a resulting dollar value per visit. The example is based on a (hypothetical) population of 100,000 patients with diabetes diagnoses; segment breakdowns are modeled on national averages where available.

**Diabetes Segments in the Patient Population**

According to the 2020 CDC National Diabetes Statistics Report, 50% of adults with diabetes had A<sub>1c</sub> values below 7, 35.5% had A<sub>1c</sub> values between 7 and 9, and 14.5% had A<sub>1c</sub> values above 9 [1]. We use this breakdown in our sample population of 100,000 patients. Next, given a lack of national data on engagement with primary care for patients with diabetes by level of glycemic control, we look at the actual health system where the intervention was deployed for engagement rates in each A<sub>1c</sub> category. Within those segments, 13.1% (6550/50,000) of the patients with A<sub>1c</sub> values below 7 were overdue for their PCP visit, as were 29.8% (10,579/35,500) of the patients with A<sub>1c</sub> values between 7 and 9, and 61.1% (8860/14,500) of the patients with A<sub>1c</sub> values above 9. Finally, 96.2% (6301/6550) of the overdue patients with A<sub>1c</sub> values below 7 did not have a future PCP visit scheduled, nor did 86.7% (9172/10,579) of the patients with A<sub>1c</sub> values between 7 and 9, nor 75.6% (6698/8860) of the patients with A<sub>1c</sub> values above 9. The number of patients from the original population of 100,000 considered unengaged with their diabetes care is 22,171 people. The calculation of the unengaged sample eligible for the intervention is summarized in Table 3.

**Table 3.** Values based on the health system where the intervention was used to describe a population of 100,000 patients with diabetes by hemoglobin A<sub>1c</sub> value who do not have a past appointment within the recommended time frame or a future visit scheduled and would therefore be eligible for the intervention.

Patient HbA <sub>1c</sub> <sup>a</sup> level	Criterion 1 (HbA <sub>1c</sub> value)	Patients (N=100,000)	And criterion 2 (appointment overdue)	Patients (n=25,989)	And criterion 3 (next appointment more than x time in future)	Total unengaged population (eligible for messages) (n=22,171)
Controlled	HbA <sub>1c</sub> <7	50,000	Last appointment >11 months	6550	3 months	6301
Moderately controlled	HbA <sub>1c</sub> ≥7 to <9	35,500	Last appointment >5 months	10,579	3 months	9172
Uncontrolled	HbA <sub>1c</sub> ≥9	14,500	Last appointment >2 months	8860	3 months	6698

<sup>a</sup>HbA<sub>1c</sub>: hemoglobin A<sub>1c</sub>.

**Identification of CPT Codes**

The national hospital data in the United States showed that approximately 67% of the patients in 2022 had commercial health insurance or were self-paying, while 20.5% had Medicare and 13.2% had Medicaid [34]. We rounded commercial health insurance coverage (which is the most lucrative for health systems) down to 66.3% in order to arrive at a total of 100%.

Applying that breakdown to the hypothetical population of 22,171 patients eligible for the intervention by A<sub>1c</sub> level allows us to look at the CPT codes each patient may be charged each year if they participate in recommended diabetes care. Then, prices from the Palmetto GBA and state Medicaid reimbursement rates are applied to calculate the dollar value associated with the recommended care for the eligible population. Because the frequency of the recommended diabetes visits varies by the A<sub>1c</sub> level, this exercise should be done

separately for each  $A_{1c}$  category. For illustrative purposes, the CPT code–based revenue potential of a diabetes well visit for the uncontrolled group ( $A_{1c}>9$ ) is described in Table 4.

Please note that the calculations for the moderately controlled and well-controlled groups are not included here but are part of

the actual analysis. For this example, the total potential annualized revenue to the health system across all 3  $A_{1c}$  categories comes to US \$44,778,974. Considering only the unengaged patients in the sample, the total potential annual reimbursement to the health system is US \$1,313,177.

**Table 4.** Potential annualized revenue from an eligible patient population with hemoglobin  $A_{1c}$  levels above 9 (uncontrolled group) based on the payer mix and the expected current procedural terminology codes that could be billed during a primary care provider visit (N=14,500).

CPT <sup>a</sup> code description	Commercial plan		Medicare		Medicaid	
	Code allocation (n=9614), n (%)	Reimbursement (USD) (Total=US \$8,986,351)	Code allocation (n=2973), n (%)	Reimbursement (USD) (Total=US \$1,852,541)	Code allocation (n=1914), n (%)	Reimbursement (USD) (Total=US \$670,742)
Physician office visit	9614 (100)	5,669,134	2973 (100)	1,225,946	1914 (100)	506,368
Hemoglobin $A_{1c}$	9614 (100)	743,081	2973 (100)	115,471	1914 (100)	68,827
Urinalysis	9614 (100)	N/A <sup>b</sup> (bundled with physician office visit)	2973 (100)	N/A (bundled with physician office visit)	1914 (100)	N/A (bundled with physician office visit)
Lipid panel	4230 (43.9)	450,869	1308 (43.9)	70,063	842 (43.9)	41,771
Complete blood count with auto-differential	9614 (100)	297,309	2973 (100)	46,200	1914 (100)	27,523
Education on self-managed blood pressure setup	6538 (68)	92,552	2022 (68)	20,014	1302 (68)	10,712
Education on self-managed blood pressure monitor	6538 (68)	1,585,158	2022 (68)	342,789	1302 (68)	N/A
Tobacco cessation	2115 (21.9)	87,894	654 (21.9)	19,007	421 (21.9)	7942
Diabetic foot examination	9614 (100)	N/A (bundled with physician office visit)	2973 (100)	N/A (bundled with physician office visit)	1914 (100)	N/A (bundled with physician office visit)
Depression	2404 (25)	60,354	743 (24.9)	13,051	479 (25)	7599

<sup>a</sup>CPT: current procedural terminology.

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

### Physician Reimbursement

Next, we calculated the expected physician reimbursement based on the CPT codes they would be able to submit for the intervention population. This provider compensation, described in Table 5, will be subtracted from the practice revenue. At this point, we assume that not all patients targeted by the intervention will participate in a PCP appointment; based on typical

engagement rates for Lirio's digital health interventions, we chose a conservative estimate of 10% appointment attendance among the total unengaged population (n=2235).

Multiplying the 5006.87 relative value units that providers can submit by the Medical Group Management Association conversion factor of US \$41.94 yields a monthly provider reimbursement of US \$17,499 or US \$209,988 per year.

**Table 5.** An estimate for provider reimbursement based on current procedural terminology codes submitted during a primary care provider appointment as predicted by comorbidity rates with diabetes.

Current procedural terminology code description	Allocation (n=2217), n (%)	Total with relative value units (n=5006.87)
Physician office visit	2217 (100)	4256.64
Hemoglobin A <sub>1c</sub> level	2217 (100)	N/A <sup>a</sup>
Urinalysis	2217 (100)	N/A
Lipid panel	975 (43.9)	N/A
Complete blood count with auto-differential	2217 (100)	N/A
Education on self-managed blood pressure setup	1508 (68)	271.36
Education on self-managed blood pressure monitor	1508 (68)	361.81
Tobacco cessation	488 (22)	117.06
Diabetic foot examination	2217 (100)	N/A
Depression	554 (24.9)	N/A

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

### Inpatient Care Reimbursement

Using the 339 in 1000 rate of hospitalization among patients with diabetes [1] and considering the 10% of the unengaged population (n=2235) that we predict to capture in the intervention, we estimate that approximately 757 patients will receive inpatient care over the next year. In our actual health system implementation, expected reimbursement per patient by payer type for an inpatient stay was provided. For purposes of the modeling exercise, we used the following values: (1) Medicare: US \$2000 based on the CMS Financial Year 2023 national adjusted operating standardized amount, non-labor-related costs [35]; (2) Medicaid: US \$1280 or 64% of the Medicare reimbursement (proportional to the reimbursement for physician office visit); and (3) commercial health plan: US \$2700 or 135% of the Medicare reimbursement. Assuming an even mix of hospitalization across payer type and A<sub>1c</sub> level and using these reimbursement values, we calculated an annual value summary from inpatient care of US \$1,796,506.

### Summary Calculation

Using the inputs described above, the final calculation using the economic impact model is as follows:

Professional practice revenue (annual, US \$1,313,177) + inpatient revenue (annual, US \$1,796,506) – provider compensation (annual, US \$209,988) = Total annual reimbursement to health system (US \$3,050,957) / Total attended PCP appointments from 100,000 patients (2217) = Dollar value per appointment (US \$1297)

The final value calculated in this example was US \$1297 per PCP appointment attended by an unengaged patient with diabetes. This value can now become the basis for pricing discussions and return on investment calculations.

## Discussion

### Principal Findings

This economic impact analysis provides a method for estimating the immediate and downstream value to a health system of the

Precision Nudging intervention targeting primary care for patients with diabetes. This model could also be applied to estimate the value for other interventions focused on connecting patients with diabetes with primary care or adapted to use for other patient populations. Being able to forecast the total economic value of an intervention is a critical prerequisite for widespread adoption [36] and will help both intervention designers and health system customers more quickly identify which tools are effective [37]. Importantly, the model takes into account the different patient risk profiles associated with various levels of glycemic control as measured by HbA<sub>1c</sub> levels. It also relies heavily on publicly available data from CMS, the CDC, and similar organizations, making it possible for other stakeholders to adapt this model to assign a value to their own interventions.

This model was developed as a pricing exercise ahead of implementing an intervention. An immediate opportunity is to populate the model with actual health system data and determine whether its predictions align with real-world performance. Precision Nudging for diabetes is currently used in the health system for whom this model was developed. Given the model's 1-year time horizon, patient claims data can be used 12 months postimplementation to verify whether the estimated impacts were realized and how the model may need to be corrected to more accurately reflect outcomes.

We anticipate the model will need to be adjusted as we learn more about the uptake and effect of interventions targeting primary care use. For example, not all patients who are targeted for a digital health intervention will take action as a result. Although in one study, 65% of adults with diabetes expressed willingness to use a digital health tool to manage their diabetes even if it had a minimal effect on their outcomes [38], the intention-behavior gap is well-documented [39] and it is well-known that digital health adoption in general lags expectations. Therefore, we recommend adjusting any economic impact estimates to reflect a portion of the population that may take action, especially if using the model as part of a pricing or sales exercise. The conservative adoption value of 10% we used

in the initial model will likely need to be edited to reflect actual engagement rates.

In fact, it is likely that a validation study will find greater uptake than the 10% estimate used to build the model. This is because the particular digital health intervention being studied meets best practices for adoption [36], such as being easy for patients to use and providing a self-evident clinical benefit (adherence to recommended provider visits). Product teams also constantly iterate and improve on digital health offerings, ostensibly improving their uptake and impact. Over time, it may be possible to quantify intervention design factors that influence uptake and use them as inputs to an economic impact model like this one. Relatedly, it will be important to reassess the intervention's impact over time as improvements are made to the intervention and how it is implemented.

Another opportunity area is to expand this type of economic modeling to digital health interventions that promote provider appointments for other chronic conditions, especially conditions where patients often have multiple comorbidities. There are likely significant downstream cost benefits to adequate primary care for patients with these conditions, and better understanding the nature of those benefits can help support health system choices on where to focus their patient support efforts.

We believe this kind of economic impact analysis will help determine the appropriate role of digital health interventions in value-based care contracts, which are increasingly common in the United States [21]. This model, along with research on leveraging primary care relationships for the care of chronic conditions, suggests that there is significant immediate and downstream value to using scalable technology to connect patients to their PCPs, such as reduced cost compared to specialist care [40], reduced mortality [41], and improved outcomes [42].

Finally, we recognize that the economic value of an intervention to a health system is only part of the story. To be successful over time, interventions must also support patient quality of life [43]; indeed, economic models focused on quantifying value to entities such as governments, employers, or payers often include quality-adjusted life years or productivity-based outcomes [23]. Although this economic value model is health system revenue-focused, it accounts for factors that matter to the patient experience. Avoiding progression of diabetes and related comorbidities and otherwise maintaining a better quality of health should positively impact patient quality of life. Ultimately, the goal is not to drive more health care utilization but rather to drive appropriate health care utilization. A future research direction is to understand patient experience consequent to interventions like Precision Nudging and ensure that the interventions deliver improvements for patients as well as systems.

### Limitations

A major limitation that undergirds the need to pressure test the model with real world data is that it is unlikely any health system

will perfectly mirror the publicly available data used to build the model. For example, we used average comorbidity rates of other health conditions for people with diabetes to estimate how frequently providers would be able to charge specific CPT codes. It is very likely that within any given health system, actual patient comorbidity rates differ from those averages. Given that part of the intent of the model is to guide pricing around primary care interventions, there is a limited acceptable margin of error for differences between the estimates and actual data. If the model is overly generous in its value calculations, it will not be accepted by health systems as a pricing tool. It also requires some effort for any health system to populate the model with their own data (eg, their patient payer mix), and it is likely some organizations will prefer not to go through the exercise. There is a tension between using average or typical data to ballpark the value of an intervention and the labor required to arrive at a more precise estimate by using actual health system data.

The inclusion of inpatient care as a value-add in the model is also a concern because it is at odds with the goals of limiting patient costs and improving quality of life. In an ideal state, primary care would help stave off hospitalizations rather than prompting them. Unfortunately, in working with unengaged patient populations, it is likely that some of them will require hospitalization subsequent to a primary visit due to unaddressed health issues. If interventions like Precision Nudging that target more regular and appropriate use of primary care are successful, over time, we hope that the value from hospitalizations in this model will need to be reduced. Finally, it is a limitation that we are unable to share the specific data used in the initial development of the model for a real health system. We have used mock data in our results that provide a similar output and offer a reasonable ballpark dollar value for the diabetes primary care visit.

### Conclusion

For digital health interventions targeting primary care to receive greater attention and be implemented in health systems, it is important to quantify the value they deliver. To our knowledge, there have been no widely accepted ways to value the economic impact of interventions that encourage appropriate use of primary care among people with diabetes. This paper offers a model based largely on publicly available data that would allow the calculation of a dollar value for a primary care visit for a patient with diabetes; this facilitates intervention pricing by vendors as well as prioritization by health systems or other customers as they consider the mix of services they use to close gaps in patient care. So much focus in digital health interventions for diabetes has been around blood glucose monitoring, education and lifestyle change, but appropriate use of primary care is a powerful tool too. Use of this model should help ensure that primary care-focused interventions receive their due recognition as effective tools to treat people with diabetes and prevent the progression of illness and its comorbidities.

## Acknowledgments

The authors wish to thank Taylor Hall, Susanne Blazek, Peter Capobianco, Matt Carson, Kelsey Ford, Cambrey Jordan, Chandra Osborn, Karilyn Housworth, and Ashley West for their support in designing the digital health intervention, developing the economic impact model, and preparing this paper.

## Conflicts of Interest

BP and AB report employment by Lirio. Neither BP nor AB have any other competing interests.

## References

1. Centers for Disease Control and Prevention (CDC). Estimates of diabetes and its burden in the United States. National Diabetes Statistics Report 2020. URL: <https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf> [accessed 2022-02-26]
2. American Diabetes Association. Economic Costs of Diabetes in the U.S. in 2017. *Diabetes Care* 2018 May;41(5):917-928 [FREE Full text] [doi: [10.2337/dci18-0007](https://doi.org/10.2337/dci18-0007)] [Medline: [29567642](https://pubmed.ncbi.nlm.nih.gov/29567642/)]
3. Lin P, Pope E, Zhou FL. Comorbidity Type and Health Care Costs in Type 2 Diabetes: A Retrospective Claims Database Analysis. *Diabetes Ther* 2018 Oct;9(5):1907-1918 [FREE Full text] [doi: [10.1007/s13300-018-0477-2](https://doi.org/10.1007/s13300-018-0477-2)] [Medline: [30097994](https://pubmed.ncbi.nlm.nih.gov/30097994/)]
4. Li R, Bilik D, Brown MB, Zhang P, Ettner SL, Ackermann RT, et al. Medical costs associated with type 2 diabetes complications and comorbidities. *Am J Manag Care* 2013 May;19(5):421-430. [Medline: [23781894](https://pubmed.ncbi.nlm.nih.gov/23781894/)]
5. Piette J, Kerr E. The Impact of Comorbid Chronic Conditions on Diabetes Care. *Diabetes Care* 2006;29(3):725-731. [doi: [10.2337/diacare.29.03.06.dc05-2078](https://doi.org/10.2337/diacare.29.03.06.dc05-2078)]
6. Choi J, Booth G, Jung HY, Lapointe-Shaw L, Tang T, Kwan JL, et al. Association of diabetes with frequency and cost of hospital admissions: a retrospective cohort study. *CMAJ Open* 2021;9(2):E406-E412 [FREE Full text] [doi: [10.9778/cmajo.20190213](https://doi.org/10.9778/cmajo.20190213)] [Medline: [33863799](https://pubmed.ncbi.nlm.nih.gov/33863799/)]
7. Robbins TD, Lim Choi Keung S, Sankar S, Randeve H, Arvanitis T. Risk factors for readmission of inpatients with diabetes: A systematic review. *J Diabetes Complications* 2019 May;33(5):398-405. [doi: [10.1016/j.jdiacomp.2019.01.004](https://doi.org/10.1016/j.jdiacomp.2019.01.004)] [Medline: [30878296](https://pubmed.ncbi.nlm.nih.gov/30878296/)]
8. Fingar K, Barrett M, Jiang H. A comparison of all-cause 7-day and 30-day readmissions. Agency for Healthcare Research and Quality. URL: <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb230-7-Day-Versus-30-Day-Readmissions.jsp> [accessed 2022-02-26]
9. Nuti LA, Lawley M, Turkcan A, Tian Z, Zhang L, Chang K, et al. No-shows to primary care appointments: subsequent acute care utilization among diabetic patients. *BMC Health Serv Res* 2012 Sep 06;12:304 [FREE Full text] [doi: [10.1186/1472-6963-12-304](https://doi.org/10.1186/1472-6963-12-304)] [Medline: [22953791](https://pubmed.ncbi.nlm.nih.gov/22953791/)]
10. Gier J. Missed appointments cost the US healthcare system \$150 B each year. *Health Care Innovation*. URL: <https://www.hcinovationgroup.com/clinical-it/article/13008175/missed-appointments-cost-the-us-healthcare-system-150b-each-year> [accessed 2022-02-26]
11. Delevry D, Ho A, Le Q. Association between processes of diabetes care and health care utilization in patients with diabetes: Evidence from a nationally representative US sample. *J Diabetes* 2021 Jan;13(1):78-88. [doi: [10.1111/1753-0407.13109](https://doi.org/10.1111/1753-0407.13109)] [Medline: [32851797](https://pubmed.ncbi.nlm.nih.gov/32851797/)]
12. Dall T, Storm M, Semilla A, Wintfeld N, O'Grady M, Narayan KMV. Value of lifestyle intervention to prevent diabetes and sequelae. *Am J Prev Med* 2015 Mar;48(3):271-280 [FREE Full text] [doi: [10.1016/j.amepre.2014.10.003](https://doi.org/10.1016/j.amepre.2014.10.003)] [Medline: [25498548](https://pubmed.ncbi.nlm.nih.gov/25498548/)]
13. Menzin J, Korn JR, Cohen J, Lobo F, Zhang B, Friedman M, et al. Relationship between glycemic control and diabetes-related hospital costs in patients with type 1 or type 2 diabetes mellitus. *J Manag Care Pharm* 2010 May;16(4):264-275. [doi: [10.18553/jmcp.2010.16.4.264](https://doi.org/10.18553/jmcp.2010.16.4.264)] [Medline: [20433217](https://pubmed.ncbi.nlm.nih.gov/20433217/)]
14. Nuti L, Turkcan A, Lawley MA, Zhang L, Sands L, McComb S. The impact of interventions on appointment and clinical outcomes for individuals with diabetes: a systematic review. *BMC Health Serv Res* 2015 Sep 02;15:355 [FREE Full text] [doi: [10.1186/s12913-015-0938-5](https://doi.org/10.1186/s12913-015-0938-5)] [Medline: [26330299](https://pubmed.ncbi.nlm.nih.gov/26330299/)]
15. 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](https://doi.org/10.1007/s00125-019-4864-7)] [Medline: [30963188](https://pubmed.ncbi.nlm.nih.gov/30963188/)]
16. Kaufman N, Mel E. Using Digital Health Technology to Prevent and Treat Diabetes. *Diabetes Technol Ther* 2020 Feb;22(S1):S63-S78. [doi: [10.1089/dia.2020.2505](https://doi.org/10.1089/dia.2020.2505)] [Medline: [32069146](https://pubmed.ncbi.nlm.nih.gov/32069146/)]
17. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do Mobile Phone Applications Improve Glycemic Control (HbA1c) in the Self-management of Diabetes? A Systematic Review, Meta-analysis, and GRADE of 14 Randomized Trials. *Diabetes Care* 2016 Nov;39(11):2089-2095. [doi: [10.2337/dc16-0346](https://doi.org/10.2337/dc16-0346)] [Medline: [27926892](https://pubmed.ncbi.nlm.nih.gov/27926892/)]
18. Castro Sweet C, Jasik CB, Diebold A, DuPuis A, Jendretzke B. Cost Savings and Reduced Health Care Utilization Associated with Participation in a Digital Diabetes Prevention Program in an Adult Workforce Population. *J Health Econ Outcomes Res* 2020;7(2):139-147 [FREE Full text] [doi: [10.36469/jheor.2020.14529](https://doi.org/10.36469/jheor.2020.14529)] [Medline: [32884964](https://pubmed.ncbi.nlm.nih.gov/32884964/)]

19. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. *J Med Internet Res* 2017 Jun 29;19(6):e232 [FREE Full text] [doi: [10.2196/jmir.7126](https://doi.org/10.2196/jmir.7126)] [Medline: [28663162](https://pubmed.ncbi.nlm.nih.gov/28663162/)]
20. Kolasa K, Kozinski G. How to Value Digital Health Interventions? A Systematic Literature Review. *International Journal of Environmental Research and Public Health* 2020;17(6):2119. [doi: [10.3390/ijerph17062119](https://doi.org/10.3390/ijerph17062119)]
21. Aitken M, Villa P, Tewary V, Anderson A. Innovation in diabetes care technology: value-based agreements. The IQVIA Institute. URL: [https://www.medicpresents.com/medicfiles/7f00\\_innovation-in-diabetes-care-technologyapril2020forweb.pdf](https://www.medicpresents.com/medicfiles/7f00_innovation-in-diabetes-care-technologyapril2020forweb.pdf) [accessed 2022-02-26]
22. Levine BJ, Close KL, Gabbay RA. Reviewing U.S. Connected Diabetes Care: The Newest Member of the Team. *Diabetes Technol Ther* 2020 Jan;22(1):1-9. [doi: [10.1089/dia.2019.0273](https://doi.org/10.1089/dia.2019.0273)] [Medline: [31483160](https://pubmed.ncbi.nlm.nih.gov/31483160/)]
23. Kwee A, Teo ZL, Ting DSW. Digital health in medicine: Important considerations in evaluating health economic analysis. *Lancet Reg Health West Pac* 2022 Jun;23:100476 [FREE Full text] [doi: [10.1016/j.lanwpc.2022.100476](https://doi.org/10.1016/j.lanwpc.2022.100476)] [Medline: [35602413](https://pubmed.ncbi.nlm.nih.gov/35602413/)]
24. Comprehensive diabetes care. NCQA/HEDIS. URL: <https://www.ncqa.org/hedis/measures/comprehensive-diabetes-care/> [accessed 2022-02-26]
25. Dulac-Arnold G, Levine N, Mankowitz DJ, Li J, Paduraru C, Goyal S, et al. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Mach Learn* 2021 Apr 22;110(9):2419-2468. [doi: [10.1007/s10994-021-05961-4](https://doi.org/10.1007/s10994-021-05961-4)]
26. Dulac-Arnold G, Mankowitz D, Hester T. Challenges of Real-World Reinforcement Learning. arXiv pre-print server 2019:1-13. [doi: [10.48550/arXiv.1904.12901](https://doi.org/10.48550/arXiv.1904.12901)]
27. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature* 2015 Feb 26;518(7540):529-533. [doi: [10.1038/nature14236](https://doi.org/10.1038/nature14236)] [Medline: [25719670](https://pubmed.ncbi.nlm.nih.gov/25719670/)]
28. Fernandez ME, Ruitter RAC, Markham CM, Kok G. Intervention Mapping: Theory- and Evidence-Based Health Promotion Program Planning: Perspective and Examples. *Front Public Health* 2019;7:209 [FREE Full text] [doi: [10.3389/fpubh.2019.00209](https://doi.org/10.3389/fpubh.2019.00209)] [Medline: [31475126](https://pubmed.ncbi.nlm.nih.gov/31475126/)]
29. 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](https://doi.org/10.1007/s12160-013-9486-6)] [Medline: [23512568](https://pubmed.ncbi.nlm.nih.gov/23512568/)]
30. Finkelstein EA, Krishnan A, Doble B. Beyond cost-effectiveness: A five-step framework for appraising the value of health technologies in Asia-Pacific. *Int J Health Plann Manage* 2020 Jan;35(1):397-408. [doi: [10.1002/hpm.2851](https://doi.org/10.1002/hpm.2851)] [Medline: [31290187](https://pubmed.ncbi.nlm.nih.gov/31290187/)]
31. Porter ME. What is value in health care? *N Engl J Med* 2010 Dec 23;363(26):2477-2481. [doi: [10.1056/NEJMp1011024](https://doi.org/10.1056/NEJMp1011024)] [Medline: [21142528](https://pubmed.ncbi.nlm.nih.gov/21142528/)]
32. American Diabetes Association Professional Practice Committee, American Diabetes Association Professional Practice Committee., Draznin B, Aroda VR, Bakris G, Benson G, et al. Comprehensive Medical Evaluation and Assessment of Comorbidities: Standards of Medical Care in Diabetes-2022. *Diabetes Care* 2022 Jan 01;45(Suppl 1):S46-S59. [doi: [10.2337/dc22-S004](https://doi.org/10.2337/dc22-S004)] [Medline: [34964869](https://pubmed.ncbi.nlm.nih.gov/34964869/)]
33. Benefits benchmarks. Medical Group Management Association Data Insights. URL: <https://tinyurl.com/6m37r6uz> [accessed 2022-02-23]
34. Breaking down US hospital payor mixes. Definitive Healthcare. URL: <https://www.definitivehc.com/resources/healthcare-insights/breaking-down-us-hospital-payor-mixes> [accessed 2022-08-02]
35. FY 2023 IPPS final rule tables. Centers for Medicare and Medicaid Services. URL: <https://www.cms.gov/medicare/acute-inpatient-pps/fy-2023-ippss-final-rule-home-page#Tables> [accessed 2022-08-02]
36. Klonoff DC, Kerr D. Overcoming Barriers to Adoption of Digital Health Tools for Diabetes. *J Diabetes Sci Technol* 2018 Jan;12(1):3-6 [FREE Full text] [doi: [10.1177/1932296817732459](https://doi.org/10.1177/1932296817732459)] [Medline: [28969460](https://pubmed.ncbi.nlm.nih.gov/28969460/)]
37. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Reply: Some considerations on digital health validation. *NPJ Digit Med* 2019;2:103 [FREE Full text] [doi: [10.1038/s41746-019-0176-z](https://doi.org/10.1038/s41746-019-0176-z)] [Medline: [31637310](https://pubmed.ncbi.nlm.nih.gov/31637310/)]
38. Oikonomidi T, Ravaud P, Cosson E, Montori V, Tran VT. Evaluation of Patient Willingness to Adopt Remote Digital Monitoring for Diabetes Management. *JAMA Netw Open* 2021 Jan 04;4(1):e2033115 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.33115](https://doi.org/10.1001/jamanetworkopen.2020.33115)] [Medline: [33439263](https://pubmed.ncbi.nlm.nih.gov/33439263/)]
39. Sheeran P, Webb TL. The Intention-Behavior Gap. *Social and Personality Psychology Compass* 2016 Sep 05;10(9):503-518. [doi: [10.1111/spc3.12265](https://doi.org/10.1111/spc3.12265)]
40. Cole E, Leighton C, Zhang Y. Distribution of visits for chronic conditions between primary care and specialist providers in Medicare shared savings accountable care organizations. *Medical Care* 2018;56(5):424-429. [doi: [10.1097/mlr.0000000000000903](https://doi.org/10.1097/mlr.0000000000000903)]
41. Basu S, Berkowitz SA, Phillips RL, Bitton A, Landon BE, Phillips RS. Association of Primary Care Physician Supply With Population Mortality in the United States, 2005-2015. *JAMA Intern Med* 2019 Apr 01;179(4):506. [doi: [10.1001/jamainternmed.2018.7624](https://doi.org/10.1001/jamainternmed.2018.7624)]

42. Mehta PP, Santiago-Torres JE, Wisely CE, Hartmann K, Makadia FA, Welker MJ, et al. Primary Care Continuity Improves Diabetic Health Outcomes: From Free Clinics to Federally Qualified Health Centers. *J Am Board Fam Med* 2016;29(3):318-324 [[FREE Full text](#)] [doi: [10.3122/jabfm.2016.03.150256](https://doi.org/10.3122/jabfm.2016.03.150256)] [Medline: [27170789](https://pubmed.ncbi.nlm.nih.gov/27170789/)]
43. Powell A, Torous J. A Patient-Centered Framework for Measuring the Economic Value of the Clinical Benefits of Digital Health Apps: Theoretical Modeling. *JMIR Ment Health* 2020 Oct 30;7(10):e18812 [[FREE Full text](#)] [doi: [10.2196/18812](https://doi.org/10.2196/18812)] [Medline: [33124995](https://pubmed.ncbi.nlm.nih.gov/33124995/)]

## Abbreviations

**CDC:** Centers for Disease Control and Prevention  
**CMS:** Centers for Medicare and Medicaid Services  
**CPT:** current procedural terminology  
**HbA<sub>1c</sub>:** hemoglobin A<sub>1c</sub>  
**PCP:** primary care provider

*Edited by A Mavragani; submitted 04.03.22; peer-reviewed by R Kadel, J Juusola; comments to author 25.08.22; revised version received 26.08.22; accepted 12.09.22; published 26.09.22.*

*Please cite as:*

*Powers B, Bucher A*

*An Economic Impact Model for Estimating the Value to Health Systems of a Digital Intervention for Diabetes Primary Care: Development and Usefulness Study*

*JMIR Form Res* 2022;6(9):e37745

URL: <https://formative.jmir.org/2022/9/e37745>

doi: [10.2196/37745](https://doi.org/10.2196/37745)

PMID: [36155985](https://pubmed.ncbi.nlm.nih.gov/36155985/)

©Brenton Powers, Amy Bucher. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 26.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Exploring Socioeconomic Status as a Global Determinant of COVID-19 Prevalence, Using Exploratory Data Analytic and Supervised Machine Learning Techniques: Algorithm Development and Validation Study

Luke Winston<sup>1</sup>, MA; Michael McCann<sup>1</sup>, PhD; George Onofrei<sup>2</sup>, PhD

<sup>1</sup>Department of Computing, Atlantic Technological University, Letterkenny, Ireland

<sup>2</sup>Department of Business, Atlantic Technological University, Letterkenny, Ireland

**Corresponding Author:**

Luke Winston, MA

Department of Computing

Atlantic Technological University

Port Road

Letterkenny, F92 FC93

Ireland

Phone: 353 862435617

Email: [L00162644@student.lyit.ie](mailto:L00162644@student.lyit.ie)

## Abstract

**Background:** The COVID-19 pandemic represents the most unprecedented global challenge in recent times. As the global community attempts to manage the pandemic in the long term, it is pivotal to understand what factors drive prevalence rates and to predict the future trajectory of the virus.

**Objective:** This study had 2 objectives. First, it tested the statistical relationship between socioeconomic status and COVID-19 prevalence. Second, it used machine learning techniques to predict cumulative COVID-19 cases in a multicountry sample of 182 countries. Taken together, these objectives will shed light on socioeconomic status as a global risk factor of the COVID-19 pandemic.

**Methods:** This research used exploratory data analysis and supervised machine learning methods. Exploratory analysis included variable distribution, variable correlations, and outlier detection. Following this, the following 3 supervised regression techniques were applied: linear regression, random forest, and adaptive boosting (AdaBoost). Results were evaluated using k-fold cross-validation and subsequently compared to analyze algorithmic suitability. The analysis involved 2 models. First, the algorithms were trained to predict 2021 COVID-19 prevalence using only 2020 reported case data. Following this, socioeconomic indicators were added as features and the algorithms were trained again. The Human Development Index (HDI) metrics of life expectancy, mean years of schooling, expected years of schooling, and gross national income were used to approximate socioeconomic status.

**Results:** All variables correlated positively with the 2021 COVID-19 prevalence, with  $R^2$  values ranging from 0.55 to 0.85. Using socioeconomic indicators, COVID-19 prevalence was predicted with a reasonable degree of accuracy. Using 2020 reported case rates as a lone predictor to predict 2021 prevalence rates, the average predictive accuracy of the algorithms was low ( $R^2=0.543$ ). When socioeconomic indicators were added alongside 2020 prevalence rates as features, the average predictive performance improved considerably ( $R^2=0.721$ ) and all error statistics decreased. Thus, adding socioeconomic indicators alongside 2020 reported case data optimized the prediction of COVID-19 prevalence to a considerable degree. Linear regression was the strongest learner with  $R^2=0.693$  on the first model and  $R^2=0.763$  on the second model, followed by random forest (0.481 and 0.722) and AdaBoost (0.454 and 0.679). Following this, the second model was retrained using a selection of additional COVID-19 risk factors (population density, median age, and vaccination uptake) instead of the HDI metrics. However, average accuracy dropped to 0.649, which highlights the value of socioeconomic status as a predictor of COVID-19 cases in the chosen sample.

**Conclusions:** The results show that socioeconomic status is an important variable to consider in future epidemiological modeling, and highlights the reality of the COVID-19 pandemic as a social phenomenon and a health care phenomenon. This paper also



puts forward new considerations about the application of statistical and machine learning techniques to understand and combat the COVID-19 pandemic.

(*JMIR Form Res* 2022;6(9):e35114) doi:[10.2196/35114](https://doi.org/10.2196/35114)

## KEYWORDS

COVID-19; machine learning; data analysis; epidemiology; human development index

## Introduction

### Background

The COVID-19 pandemic represents the most unprecedented global challenge in recent times. Originally identified in the city of Wuhan, China, the SARS-CoV-2 virus spread across the world, and the situation escalated into an international emergency. Despite widescale containment efforts in 2020, as well as the largest vaccine rollout in history [1], the pandemic continued to challenge the global community in 2021. Research is being conducted to analyze the trajectory of the virus, and to understand why particular populations or countries have been more severely impacted than others [2,3]. This has been supported by increases in data availability, which has enabled researchers to investigate a large range of potential COVID-19 risk factors. These risk factors can be categorized as clinical or nonclinical. Clinical risk factors include obesity [4-6], diabetes [7,8], and smoking [9]. Examples of nonclinical risk factors are cultural differences [10], government containment measures [11], vaccination attitudes [12], and socioeconomic status [13-15].

This paper focuses on the nonclinical risk factor of socioeconomic status as a determinant of COVID-19 prevalence. To provide a reliable empirical metric for socioeconomic status, the Human Development Index (HDI) of the United Nations Development Programme (UNDP) was selected. The HDI calculates the overall socioeconomic status or “well-being” of inhabitants in a country by aggregating its life expectancy, education, and per capita income metrics [16]. It has been applied successfully in previous epidemiological research to map prevalence rates of various diseases [17-20]. Despite its popularity in statistical analysis, the HDI has not yet been widely applied in machine learning COVID-19 modeling. This presents an opportunity to apply statistical and machine learning techniques to examine whether the HDI can be used to accurately predict prevalence rates of COVID-19.

### Related Work

#### *Socioeconomic Status in Health Research*

Pandemics are as much a social problem as a health care problem [21]. As such, socioeconomic status is an important determinant to consider in pandemic research. The term socioeconomic status is an umbrella term used to describe empirically measurable social or economic factors, such as social class, education, income, and health [22,23]. These factors are applied in a variety of ways to investigate or control their effects on given outcomes, such as health outcomes, and have consistently been found to be statistically significant [24-26]. In terms of health outcomes, higher socioeconomic status has typically been associated with better health. Conversely, lower

socioeconomic status is associated with poorer health outcomes [27]. In the literature, lower socioeconomic status has been associated with higher rates of illnesses, such as osteoarthritis, chronic diseases, hypertension, and cervical cancer [28,29].

In relation to COVID-19, socioeconomic status has also been associated with higher prevalence and more severe outcomes. In the United States, the Distressed Communities Index has been used to analyze the impact of socioeconomic status on COVID cases and mortality [30]. Results from this study indicated that lower education and racial differences were associated with poorer COVID-19 outcomes. Another study argued that lower socioeconomic populations are more likely to live in overcrowded accommodation and have less access to outdoor space, making them more vulnerable to COVID-19 infection [31]. Evidently, socioeconomic status is an important determinant of COVID-19 outcomes, which can uncover how the virus affects particular populations.

### HDI

The HDI is a composite measure of overall socioeconomic status at the national level, which is annually calculated by the UNDP. The HDI indices include life expectancy, expected years of schooling, mean years of schooling, and gross national income (GNI). Calculating a country’s HDI for a given year requires 2 steps. First, values from each of the 4 indices are normalized to an index value between 0 and 1. Maximum and minimum limits for each metric are set by the UNDP. Using the actual value, maximum value, and minimum value, the dimension index can be calculated with the following formula:

$$\text{Dimension index} = (\text{actual value} - \text{minimum value}) / (\text{maximum value} - \text{minimum value})$$

Second, once each individual dimension has been calculated, the equally weighted mean is calculated to provide the overall HDI score of a country [32].

The HDI has been used in health research to analyze both the prevalence rates and mortality rates of specific diseases, which helps to identify disparities in terms of outcome within a country or between countries. It has been applied to understand a range of epidemiological research problems, such as malaria [17], various cancer distributions [19,33,34], hypertension [20], *Blastocystis* parasites [35], and dental health [36]. To provide a specific example, research investigating the relationship between the HDI and thyroid cancer suggested that although higher HDI countries have a higher prevalence of the disease, lower HDI countries have higher mortality rates [34].

The HDI has also been applied to analyze the ongoing COVID-19 pandemic, generating important insights about the disproportionate impact of the pandemic cross-nationally. For example, a study analyzing the HDI and COVID-19 mortality

reported that countries with high HDI scores recorded higher COVID-19 mortality rates [13]. Another study reported significant correlations between the HDI scores of 166 countries and their confirmed cases on March 27, 2020 [14]. Elsewhere, a study focusing on municipal differences in COVID-19 impact in Brazil (using a recalibrated index to analyze municipal differences rather than national differences) found that municipalities with high HDI scores had the highest COVID-19 incidence rate and mortality per 100,000 population as of May 2020 [15]. The index has therefore been recognized as a valuable framework in COVID-19 research.

### **Multicountry COVID-19 Research**

Multicountry COVID-19 research is important for the following 2 reasons: (1) the ability to identify country-specific points of interest, and (2) the ability to uncover common trends or risk factors across countries. In a study of lockdown-associated mental health problems in Egypt, Pakistan, India, Ghana, and the Philippines, it was reported that although lockdowns negatively affected the mental health of respondents in each country, they did so in different ways. For example, respondents from the Philippines coped with lockdowns by increasing self-destructive behaviors, while those from Pakistan sought comfort in religion. Respondents from the 3 remaining countries tended to accept the lockdowns [37]. A similar study in a larger sample of 101 countries analyzed the loneliness and social isolation associated with the pandemic [38]. Other studies have been conducted to analyze cross-national vaccination attitudes [39], the success of containment measures [11,40], and cultural behaviors that impacted cross-national COVID-19 mortality rates [10]. Therefore, multicountry COVID-19 research helps to identify “global risk factors” relating to the pandemic, subsequently aiding evidence-based public health interventions [38]. It also opens up new research questions as to why certain populations behaved or were impacted a certain way during the pandemic.

### **Modeling Outbreaks Using Machine Learning**

When modeling outbreaks, a popular method in epidemiology is the susceptible, infected, recovered (SIR) approach. The SIR approach simplifies the transmission dynamics of infectious diseases by dividing the population into groupings of susceptible, infected, and recovered individuals and analyzes the interaction between these groups over the course of an outbreak. This method has also been deployed to analyze the COVID-19 pandemic [41,42]. However, SIR modeling assumes that complete herd immunity is possible through infection [43], which limits its efficacy in COVID-19 research. It is not yet understood if COVID-19 herd immunity is achievable due to the complex nature of the virus, the questionable long-term efficacy of available vaccines, the emergence of new variants, and the cases of reinfection [44]. Subsequently, the predictive benefits of machine learning may yield better results in relation to this pandemic.

Advancements in machine learning have enabled epidemiological researchers to use a robust data-driven approach facilitated by high-precision algorithms. This has helped to process ever-increasing volumes of data, and to analyze a wider range of factors that impact patient health outcomes [45,46].

For example, naïve Bayes, logistic regression, random forest, and artificial neural network models have been developed to predict hypotension in patients after receiving an anesthetic [47]. Elsewhere, gated recurrent unit neural networks have been designed to identify individuals at risk of in-hospital mortality. This model allows practitioners to map the probability of death longitudinally, and to provide targeted interventions based on the model predictions [48].

Another advantage of machine learning in epidemiology is that it can predict and map disease occurrences and health outcomes in situations where data are limited [49]. Specifically, boosted regression tree models have been used to analyze environmental factors that affect the transmission of diseases, such as dengue fever, Ebola, Crimean-Congo hemorrhagic fever, and Zika virus [50-53]. Another type of machine learning model, the Ensemble Adjustment Kalman Filter, has been used to forecast seasonal outbreaks of influenza [54]. Additionally, several retrospective forecasting studies have been conducted to reconstruct past pandemics, including Ebola, West Nile Virus, and Respiratory Syncytial Virus, by mapping their transmission patterns [55-57].

Regarding COVID-19, epidemiological research using machine learning is emerging in the literature at pace. Generally, studies have involved the design of one or more machine learning models to predict COVID-19 case prevalence [11,58,59], severity [60,61], and mortality/risk of mortality [62,63]. In 1 study, 5 non-time series supervised learning models using random forest and AdaBoost regression were trained to predict the confirmed infection growth (the 14-day growth of the cumulative number of reported COVID-19 cases) of COVID-19 in 114 countries, using nonpharmaceutical containment measures and cultural dimensions as features. Results indicated that confirmed infection growth was predicted to a considerable degree with moderate to high  $R^2$  values ( $>0.50$ ) [11]. Lastly, a systematic review of machine learning techniques in the prediction of COVID-19 cases found that  $R^2$  values ranged between 0.64 and 1, suggesting that machine learning is a highly valuable method for predicting COVID-19 prevalence, which could support policy makers in shaping future interventions [64].

### **Description of the Study**

This study analyzed the statistical relationship between HDI scores and cumulative COVID-19 cases (total recorded cases up to December 31, 2021) in a sample of 182 countries. It then attempted to predict 2021 COVID-19 cumulative cases in the sample using the previous year’s cumulative cases (total recorded cases up to December 31, 2020) and HDI scores. Cumulative cases per million of the population was selected as it provides the number of reported infections proportionate to the population size. Crude rate metrics, such as cases per million, are the most effective for multicountry samples [65]. For example, Afghanistan and Albania reported a similar absolute number of COVID-19 cases in 2020, with values of 51,526 and 58,316, respectively. However, Afghanistan’s cases per million was 1324, while Albania’s was 20,264. This shows the viral prevalence relative to both populations and indicates that Albania actually had higher case rates in 2020.

To measure socioeconomic status, the HDI indices of life expectancy, expected years of schooling, mean years of schooling, and GNI were used. For the purposes of this study, individual metrics were selected rather than the aggregated HDI value. This approach was used because aggregation can lose important information in the data, which can lead to less accurate predictions [66].

Two predictive models were designed using the open-source integrated development environment Jupyter Notebook, which is compatible with Python programming language. Each model was trained using the following 3 supervised learning regression algorithms: basic linear regression, random forest, and AdaBoost. All algorithms were evaluated using k-fold cross-validation and then compared by calculating their  $R^2$  scores and error statistics. The first model attempted to predict 2021 COVID-19 prevalence using 2020 case numbers to establish a baseline for the performance of the second model. The second model included 2020 case numbers and each country's life expectancy, expected years of schooling, mean years of schooling, and GNI metrics. Due to the uneven progress of the pandemic on a country-by-country basis, this study focused on cross-sectional data rather than time-series data. All data for this study are secondary and publicly available, highlighting the commendable global effort to collect and share data concerning the pandemic.

## Methods

### Data Preprocessing

COVID-19 case data were downloaded from the COVID-19 OurWorldInData database [65], which in turn retrieves data from the John Hopkins Center for Systems Science and

Engineering Data Repository. The OurWorldInData database contains comprehensive COVID-19 metrics for 190 countries, including infection rates, hospitalization numbers, mortality rates, and vaccination uptake figures. Data are uploaded daily, which allows users to track the evolution of the pandemic with up-to-date statistics. This research required each country's "cases per million" figure for December 31, 2020, and the same metric for December 30, 2021. HDI data were extracted from the 2020 Human Development Report Data Center [67]. The report provides each country's overall HDI score and the score for each individual metric.

These data sets were combined so that each observation (country) contained the following metrics: (1) life expectancy, (2) expected years of schooling, (3) mean years of schooling, (4) GNI, (5) COVID-19 cases per million in 2020 (January 1-December 31), and (6) COVID-19 cases per million in 2021 (January 1-December 31).

Countries with missing data were omitted; therefore, the final data set contained data for 182 countries. It was then imported to Jupyter and converted into dataframe format (see [Table 1](#)) to begin analysis.

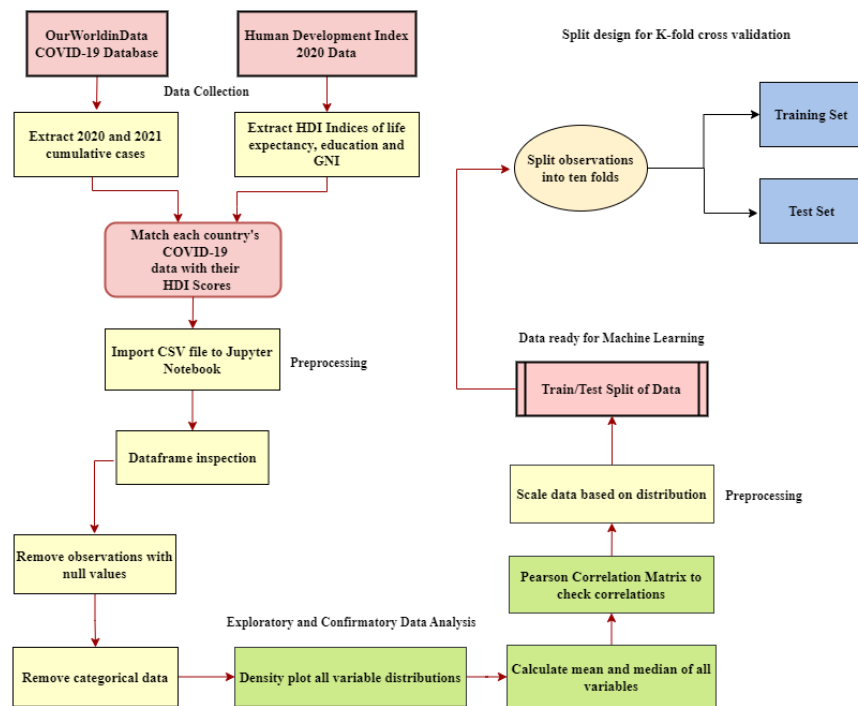
Following this, exploratory data analysis was conducted to explore the distribution of the data and the statistical relationships between the variables. A data scaling method was then selected depending on the distribution of the data. Data scaling is important in machine learning modeling as it prevents measurement differences from negatively affecting the final results [68]. The interquartile range was then calculated to identify outliers in the target variable (2021 COVID-19 cases).

[Figure 1](#) summarizes the workflow for this study, from data preprocessing to model design and exploratory data analysis.

**Table 1.** Sample of the data set using Human Development Index metrics and COVID-19 cases.

Country	Life expectancy	Expected years of schooling	Mean years of schooling	Gross national income per capita (US\$)	Cases 2020 (per million)	Cases 2021 (per million)
Afghanistan	64.8	10.2	3.9	2239	1323.612	3968.427
Albania	78.6	14.7	10.1	13,998	20,264.091	73,173.975
Algeria	76.9	14.6	8.0	11,174	2271.554	4895.753
Andorra	81.9	13.3	10.5	56,000	104,173.947	306,900.742
Angola	61.2	11.8	5.2	6104	534.073	2404.489

**Figure 1.** A flowchart illustrating the data pipeline, from the collection of COVID-19 and Human Development Index (HDI) data to the cross-validation training and testing process. In addition to designing the predictive models, exploratory data analysis was also conducted to identify trends in the data set. GNI: gross national income.



## Machine Learning Algorithm Selection

Supervised machine learning models are trained to make predictions by learning from a data set where the value of the output (dependent variable) is known for each observation. Supervised machine learning produces decisions or “outputs” based on input data during the training process. Implementing different supervised algorithms on a set of data allows for the results to be compared and for the best fitting model to be identified [69,70]. Evaluating a supervised learning model requires robust validation measures [71]. These can be calculated using a variety of accuracy and error metrics, such as the coefficient of determination ( $R^2$ ), mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), or max error. This research compared the performances of linear regression, random forest, and AdaBoost supervised techniques.

### Linear Regression

Linear regression is one of the most common machine learning algorithms [72]. Regression in machine learning differs from traditional statistical regression as it partitions the data set into a training set and a test set. Using the input and output data from the training set, algorithms attempt to predict output data in the test set using input data only. This process indicates how accurately a model can make predictions on new data. Linear regression is calculated as follows:

$$y = a_0 + a_1x + \epsilon$$

where  $y$  is the target variable (output),  $x$  is the predictor variable (input),  $a_0$  is the intercept,  $a_1$  is the coefficient, and  $\epsilon$  is random error.

### Random Forest

Random forest is an ensemble of decision tree algorithms that can be used for either classification or regression problems. It is based on the concept of bagging or bootstrap aggregation, which creates an ensemble of learner trees [73]. Each learner tree ( $K$ ) is trained on separate samples drawn from the original data set (input vector  $x$ ), and the overall prediction is obtained by calculating the mean of  $K$  regression trees as follows:



Random forest is beneficial for reducing model variance compared to individual decision trees. It also helps to prevent model overfitting (when a model fits too closely to training data and poorly to test data) [74].

### AdaBoost

AdaBoost or adaptive boosting is a sequential ensemble technique that is based on the principle of developing several weak learners using different training subsets drawn randomly from the original training data set. Using this technique, the training algorithm begins with 1 decision tree, identifies the observations with the highest error, and adds more weight to these. The weights are recalculated after every iteration so that incorrectly classified observations by the previous decision tree receive higher weights [75]. Using Python programming language, the number of trees that the algorithm will deploy can be chosen, with the default set at 50 iterations.

### Model Design and Evaluation

Two feature models were created (Feature Model 1 and Feature Model 2). Feature Model 1 was trained to predict 2021 COVID-19 prevalence using 2020 cases only. Feature Model 2 was trained to predict 2021 COVID-19 prevalence using 2020

case data as well as life expectancy, expected years of schooling, mean years of schooling, and GNI per capita. Each feature model was trained using linear regression, random forest, and AdaBoost techniques. Hyperparameters were set for each algorithm, and results were evaluated using a 10-fold (k=10) k-fold cross-validation.

**Model Hyperparameters and Validation**

Rather than partitioning the data into training and test sets using the train/test split, this research used k-fold cross-validation. K-fold cross-validation has a single parameter called *k* that represents the number of subsets or “folds” that a data set will be split into, which the user selects. As shown in Figure 2, each fold uses a different grouping of data as the test set, and the process is then repeated *k* number of times (for example, 5 times in Figure 2). It is evaluated by the cross-validation score, which is the mean of all scores from each k-fold subset. K-fold cross-validation provides a more generalizable and less biased performance estimate when working with smaller data sets [76,77]. This is because it maximizes the number of observations that can be used for both training and testing. In other words, a model using cross-validation does not depend on a single train/test split.

Using sklearn, the mean cross-validation score defaults to the scoring metric for the specific algorithm being cross-validated. For each algorithm in this study, the default scoring metric was the coefficient of determination ( $R^2$ ). Therefore, the mean cross-validation score computed was the average  $R^2$  for each algorithm across all k-folds.  $R^2$  represents the goodness of fit of a regression model and explains how much variance in the dependent variable can be explained by one or more independent variables. It is calculated by dividing the residual sum of squares by the total sum of squares and subtracting the derivation from 1, as follows:

$$R^2 = 1 - (\text{residual sum of squares} / \text{total sum of squares})$$

$R^2$  was the primary measure under observation in this study. In machine learning,  $R^2$  is the most informative validation measure with the least interpretive limitations [78].

Table 2 presents the hyperparameters unique to each algorithm. A 10-fold validation was selected for the k-fold cross-validation, which is a generally recommended number of subsets to apply [76,77].

Alongside  $R^2$ , 4 error metrics were also calculated to assess performance. First, MAE provides the average of the absolute error between the predicted values and true values. It is calculated as follows:

$$\frac{1}{n} \sum |y_i - x_i|$$

where  $y_i$  is the prediction value,  $x_i$  is the actual value, and  $n$  is the number of observations.

Second, MSE measures the average squared difference between the predicted values and true values. It is calculated as follows:

$$\frac{1}{n} \sum (Y_i - \hat{Y}_i)^2$$

where  $n$  is the number of data points,  $Y_i$  is the actual value, and  $\hat{Y}_i$  is the predicted value.

Third, RMSE calculates the square root of the mean of squared errors of a model. It is calculated as follows:

$$\sqrt{\frac{1}{N} \sum (x_i - \hat{x}_i)^2}$$

where  $i$  is the variable  $i$ ,  $N$  is the number of data points,  $x_i$  is the actual value, and  $\hat{x}_i$  is the predicted value.

Finally, max error computes the maximum residual error, which captures the worst case error between the predicted value and the true value. It is calculated as follows:

$$\max |y_i - \hat{y}_i|$$

where  $\hat{y}$  is the predicted value of the  $i$ -th sample, and  $y_i$  is the corresponding true value.

**Figure 2.** An example of the 5-fold k cross-validation method where k=5. The overall accuracy score is calculated as the mean value of each fold’s accuracy score.



**Table 2.** Supervised learning model hyperparameters using cross-validation.

Algorithm	Hyperparameters
Basic linear regression	Folds: 10; random state: 1
Random forest	Folds: 10; random state: 1; estimators: 100
AdaBoost	Partitions: 10; estimators: 50; random state: 0

## Results

### Exploratory Data Analysis

Exploratory data analysis was carried out to identify and visualize trends in the data, and to statistically analyze the variables. In 2020, the mean number of COVID-19 cases per million in the sample was 15,880.41, with a median of 6822.98. In 2021, the mean number of COVID-19 cases per million was 64,479.58, with a median of 50,764.73. [Table 3](#) presents the key descriptive statistics of all variables in the study.

Distplots were created to inspect the distribution of all variables. The resulting plots showed that all variables, with the exception of expected years of schooling, were skewed in the sample. The distribution of 2021 COVID-19 prevalence was positively

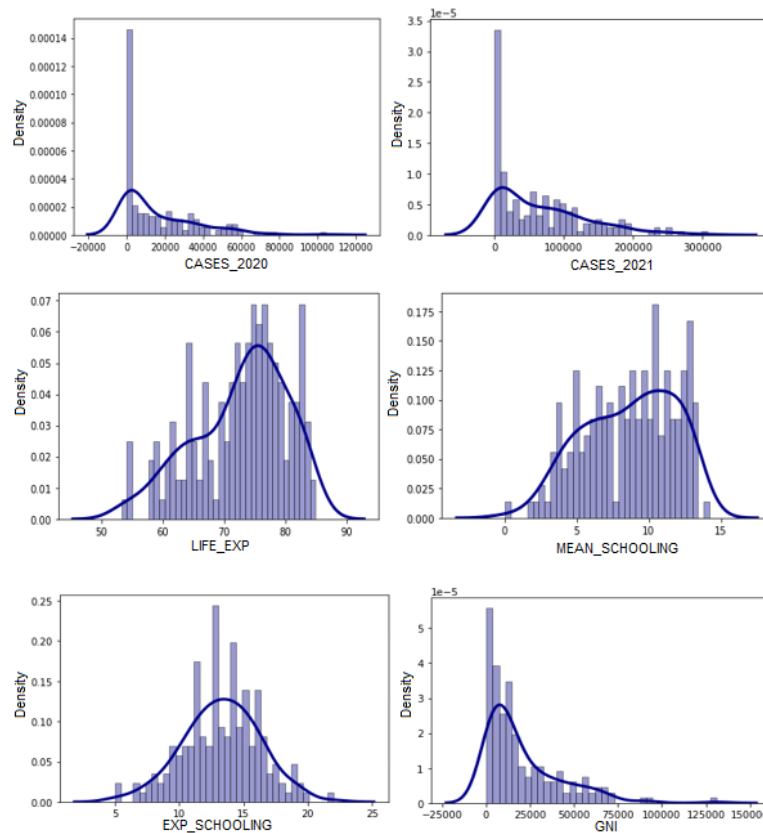
skewed in the sample (see [Figure 3](#)). Calculation of the interquartile range revealed that 4 countries (Andorra, Montenegro, Serbia, and Seychelles) were statistical outliers, which had recorded unusually high rates of COVID-19 (>250,000 per million population). The Seychelles recorded the highest prevalence with 217,096.35 cases per million.

To investigate the statistical relationship between the features and the target variable, a Pearson correlation matrix was implemented (see [Figure 4](#)). All chosen features correlated statistically with 2021 COVID-19 prevalence, with R values ranging from 0.55 to 0.85. Moreover, 2020 COVID-19 cases had the strongest correlation with 2021 case data (R=0.85), followed by mean years of schooling (R=0.66), life expectancy (R=0.61), expected years of schooling (R=0.58), and GNI (R=0.55).

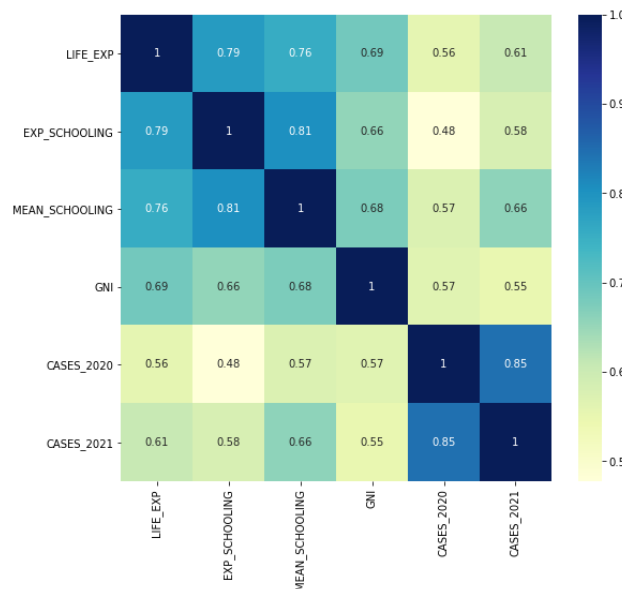
**Table 3.** Statistical measurements (mean and median) of all variables in the study.

Variable	Mean value	Median value
2020 COVID-19 cases per million	15,880.41	6822.98
2021 COVID-19 cases per million	64,479.58	50,764.73
Life expectancy	72.72	74.20
Expected years of schooling	13.31	13.15
Mean years of schooling	8.63	8.95
Gross national income per capita (US\$)	20,453.40	13,112.50

**Figure 3.** A series of density plots illustrating the distribution of each variable under observation (the target variable). The target variable 2021 COVID-19 cases per million is right-skewed in the sample. Expected years of schooling is the only variable with a normal distribution in the sample. CASES\_2020: 2020 COVID-19 cases per million; CASES\_2021: 2021 COVID-19 cases per million; EXP\_SCHOOLING: expected years of schooling; GNI: gross national income per capita; LIFE\_EXP: life expectancy; MEAN\_SCHOOLING: mean years of schooling.



**Figure 4.** Pearson correlation matrix mapping the correlation between all variables. Results show that all features have a statistical correlation with 2021 COVID-19 cases. CASES\_2020: 2020 COVID-19 cases per million; CASES\_2021: 2021 COVID-19 cases per million; EXP\_SCHOOLING: expected years of schooling; GNI: gross national income per capita; LIFE\_EXP: life expectancy; MEAN\_SCHOOLING: mean years of schooling.



**Supervised Learning Models**

Tables 4 and 5 summarize the performances of all regression algorithms in both feature models, while Figure 5 visualizes their performances. Feature Model 1 was trained to predict 2021 COVID-19 cases per million using 2020 cases per million

(n=182). Feature Model 2 was trained to predict 2021 COVID-19 cases per million using 2020 cases per million as well as life expectancy, mean years of schooling, expected years of schooling, and GNI (n=182). Both data sets were divided into 10 folds for cross-validation (k=10).

In Feature Model 1, linear regression was the most accurate learner with a mean  $R^2$  of 0.693, followed by random forest (0.481) and then AdaBoost (0.454). The variation in performance was considerable, with a 23.9% difference between the most precise and least precise algorithms. In Feature Model 2, the basic linear regression model was also the strongest learner ( $R^2=0.762$ ), followed by random forest (0.722) and AdaBoost (0.679). The MAE, MSE, RMSE, and max error statistics of the algorithms were all lower in Feature Model 2 than in Feature Model 1. Feature Model 2 also exhibited closer performances between the algorithms than Feature Model 1, with the strongest learner being 8.4% more precise than the least.

Although it was the best learner on the data in both models, linear regression showed the least improvement with the inclusion of socioeconomic indicators in Feature Model 2 ( $R^2$  improved by 7%). Additionally, its error statistics did not improve as significantly as those of random forest or AdaBoost. For example, the MAE of linear regression decreased by 0.009 (0.079 in Feature Model 1 and 0.070 in Feature Model 2) compared to decreases of 0.026 in random forest and 0.014 in AdaBoost.

Tables 6 and 7 outline the performance accuracy of each individual fold. The widely varying  $R^2$  scores indicate that the cross-validation approach used in this study yielded the most reliable results.

**Table 4.** Evaluation of Feature Model 1 using linear regression, random forest, and AdaBoost.

Evaluation measure	Linear regression <sup>a</sup>	Random forest <sup>a</sup>	AdaBoost <sup>a</sup>
$R^2$	0.693	0.481	0.454
MAE <sup>b</sup>	0.079	0.096	0.104
MSE <sup>c</sup>	0.014	0.021	0.020
RMSE <sup>d</sup>	0.117	0.143	0.142
Max error	0.315	0.359	0.355

<sup>a</sup>All results were evaluated using k-fold cross-validation (k=10).

<sup>b</sup>MAE: mean absolute error.

<sup>c</sup>MSE: mean squared error.

<sup>d</sup>RMSE: root mean squared error.

**Table 5.** Evaluation of Feature Model 2 using linear regression, random forest, and AdaBoost.

Evaluation measure	Linear regression <sup>a</sup>	Random forest <sup>a</sup>	AdaBoost <sup>a</sup>
$R^2$	0.763	0.722	0.679
MAE <sup>b</sup>	0.070	0.070	0.090
MSE <sup>c</sup>	0.011	0.013	0.015
RMSE <sup>d</sup>	0.107	0.114	0.124
Max error	0.265	0.308	0.300

<sup>a</sup>All results were evaluated using k-fold cross-validation (k=10).

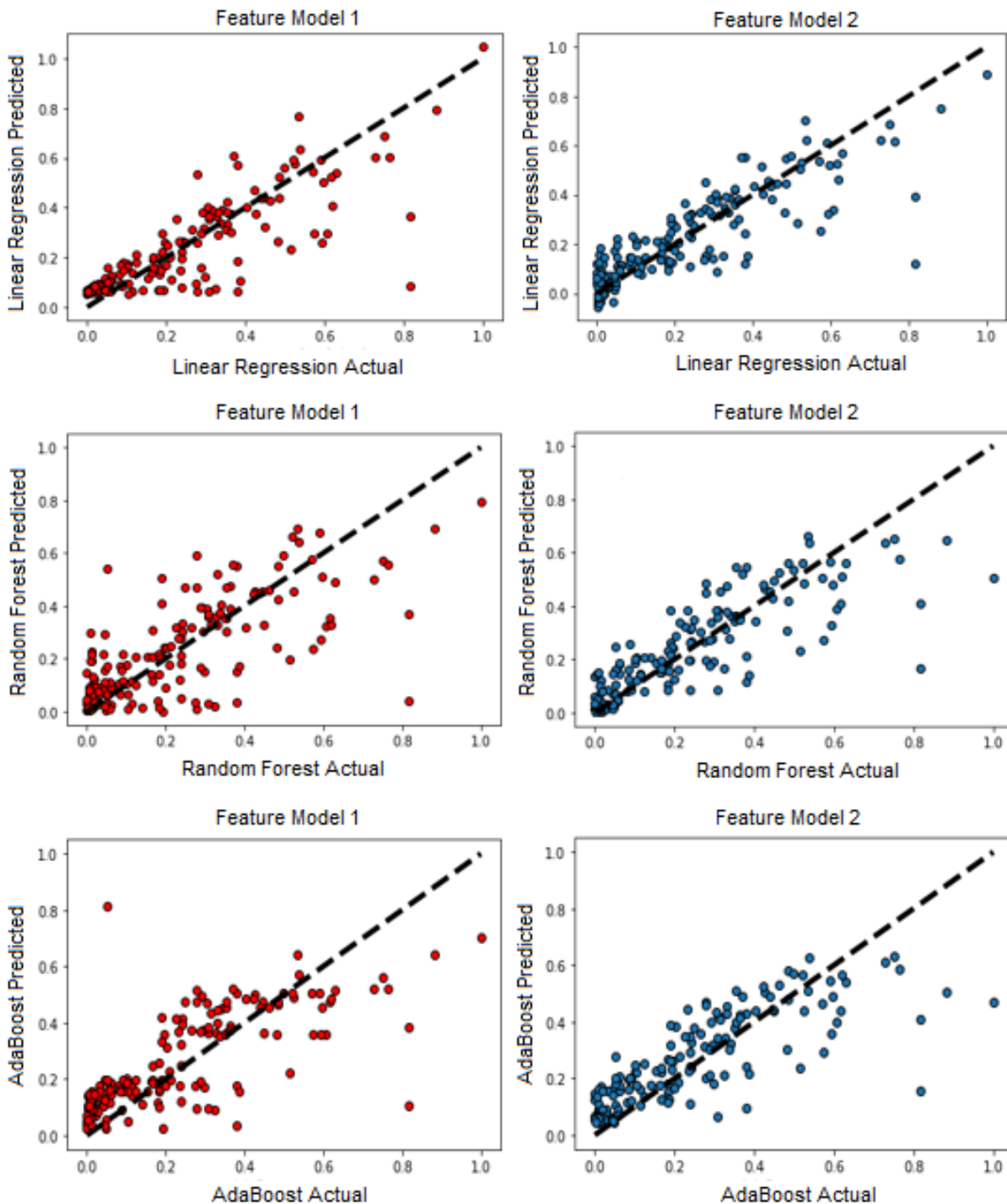
<sup>b</sup>MAE: mean absolute error.

<sup>c</sup>MSE: mean squared error.

<sup>d</sup>RMSE: root mean squared error.



**Figure 5.** A series of subplots showing the predictive performances of the linear regression, random forest, and AdaBoost algorithms in both Feature Models 1 and 2. Each observation represents a prediction of 2021 COVID-19 cumulative cases per million, with the regression line being the true value. With the addition of Human Development Index metrics, the linear regression algorithm improved from  $R^2=0.693$  to  $0.763$ . The random forest algorithm improved from  $R^2=0.481$  to  $0.722$ . The AdaBoost algorithm improved from  $R^2=0.454$  to  $0.679$ . Data points were calculated using `cross_val_predict`, which shows the predicted output from each test set within each k fold.



**Table 6.** Accuracy for each algorithm's individual fold (k=10) in Feature Model 1.

Iteration	Linear regression	Random forest	AdaBoost
Fold 1	0.877	0.799	0.759
Fold 2	0.768	0.687	0.342
Fold 3	0.657	0.464	0.584
Fold 4	0.803	0.530	0.629
Fold 5	0.747	0.153	-0.696
Fold 6	0.733	0.553	0.766
Fold 7	0.804	0.628	0.652
Fold 8	0.035	-0.287	0.083
Fold 9	0.767	0.627	0.696
Fold 10	0.742	0.657	0.722

**Table 7.** Accuracy for each algorithm's individual fold (k=10) in Feature Model 2.

Iteration	Linear regression	Random forest	AdaBoost
Fold 1	0.774	0.796	0.679
Fold 2	0.595	0.457	0.485
Fold 3	0.946	0.907	0.882
Fold 4	0.602	0.622	0.551
Fold 5	0.833	0.869	0.824
Fold 6	0.780	0.776	0.720
Fold 7	0.627	0.636	0.626
Fold 8	0.850	0.659	0.536
Fold 9	0.780	0.794	0.851
Fold 10	0.844	0.594	0.629

## Discussion

### Principal Findings

Results from exploratory data analysis yielded a number of interesting insights. First, the positively skewed distribution of 2021 COVID-19 cases resulted in a mean greater than the median in the sample. In the 182 countries sampled, COVID-19 prevalence was asymmetrical and revealed that a minority of countries recorded very high case numbers. Second, the distribution of 2020 COVID-19 cases was positively skewed and similar visually to the 2021 distribution. This shows that the trajectory of the virus in the sample was relatively consistent in 2020 and 2021 in terms of cumulative reported cases. Third, the 4 outlier countries identified shared an interesting pattern; all had higher than average life expectancy, mean years of schooling, and GNI compared with the means in the sample. This indicates that the outliers can be considered above average socioeconomically. Finally, all HDI metrics correlated positively with COVID-19 cases per million, which points to an important statistical relationship between socioeconomic status and COVID-19 prevalence. Education (expected/mean years) shared the highest correlation, followed by life expectancy and then GNI. This correlation is noteworthy and highlights the unique nature of the COVID-19 pandemic. Typically, lower

socioeconomic status is associated with poorer health outcomes, but the results from this study suggest that countries with higher socioeconomic status recorded higher rates of COVID-19 in 2021. This could be because more developed countries tend to have older populations, as well as higher prevalence of known COVID-19 clinical risk factors, such as diabetes and cardiovascular disease [79].

The results from machine learning analysis suggest that 2021 COVID-19 prevalence could be predicted with a reasonable degree of accuracy using the previous year's prevalence rates and the socioeconomic indicators of life expectancy, mean years of schooling, expected years of schooling, and GNI per capita. With socioeconomic indicators included, the  $R^2$  of each learning algorithm was higher than that when trained on only 2020 COVID-19 data, and the error statistics were lower. Including the HDI indices as predictors alongside the previous year's COVID-19 cases in each country improved the predictive accuracy of 2021 cases by an average of 18% across the 3 chosen algorithms. Given that predictive algorithms can struggle with smaller data sets [59], the results of this study (n=182) are noteworthy.

The linear regression algorithm was the strongest learner on the data, but also showed the least improvement (7% increase in

mean cross-validation) once the HDI metrics were added. Given that the other algorithms improved considerably when HDI indices were added, this result represents an interesting outlier. The varying performances between the algorithms may be due to the statistically linear relationships between the variables (as discovered in the Pearson correlation matrix in Figure 4). Despite the strong correlation between 2021 COVID-19 cumulative cases per million and case data from the previous year ( $R=0.84$ ), Feature Model 1 did not make accurate predictions using random forest or AdaBoost. Unlike linear regression models, which excel at fitting to data where linear correlation exists, decision tree algorithms like random forest and AdaBoost may perform more effectively with nonlinear data [80,81]. Lastly, the widely varying performance of each k-fold iteration justified the use of cross-validation to evaluate the models. In Feature Model 2, for example, the highest scoring fold of the linear regression algorithm had a result of 94.6, a highly accurate  $R^2$  result. However, the lowest scoring fold had an  $R^2$  of 59.5. The cross-validation  $R^2$  score of 76.3 was therefore the most reliable score for the data set.

### Follow-Up Analysis

Following the primary analysis, 4 follow-up analyses were conducted. First, Feature Model 2 was trained again without 2020 COVID-19 case data as a feature to analyze how well the HDI metrics could predict COVID-19 cases alone. Without the previous year's case data, the accuracy was low ( $R^2=0.438$  for the best performing algorithm, which was again linear regression). This result highlights the significant importance of 2020 case data in predicting the following year's COVID-19 prevalence. Second, Feature Model 2 was trained again using 1 HDI metric at a time to analyze which was the most important for the prediction of COVID-19 cases. The results showed that expected years of schooling and mean years of schooling had the highest scores ( $R^2=0.755$  for each), followed by life expectancy ( $R^2=0.739$ ) and then GNI ( $R^2=0.712$ ). This suggests that education was the most predictive socioeconomic indicator (the education HDI metrics were also the most statistically correlative). However, the results also showed that using all HDI indices is more effective than using them separately for COVID-19 case prediction in this data set. The third follow-up

experiment removed the 4 previously identified outlier countries (Andorra, Montenegro, Serbia, and Seychelles) from the data set and implemented both feature models again, using the same cross-validation method as the initial analysis. This yielded interesting results (see Tables 8 and 9). Most notably, random forest became the strongest learner in Feature Model 2 ( $R^2=0.777$ ). Despite being generally less sensitive to outliers [82], random forest benefitted from outlier removal in this data set. Removing outliers also reduced the gap in performance between the algorithms. With outliers included, Feature Model 1 displayed a 23.9% difference between the best and worst performing algorithms, and with outliers removed, this difference reduced to 19.5%. This reduction was more apparent in Feature Model 2, with just a 2.1% difference between the best and worst performing algorithms with outliers removed (compared with an 8.4% difference in the original sample with outliers included). However, the results indicated that removing the outliers did not significantly improve overall predictive accuracy.

The fourth follow-up experiment sought to compare socioeconomic status as a COVID-19 predictor with a selection of other COVID-19 risk factors. Subsequently, each country's median age, population density (individuals per square kilometer), and percentage of vaccinated individuals were sourced and added to the data set. Each of these variables has been shown to predict COVID-19 prevalence in certain samples [83-85]. Most of the required data were also available in the OurWorldInData database, though a small number of entries had to be sourced from Worldometers and IndexMundi [86,87].

When Feature Model 2 was trained again using these new metrics alongside 2020 case data, predictive accuracy dropped to an average of 0.649 across all 3 algorithms. Using these new features, the most accurate algorithm was 10% less accurate than the most accurate learner in the model with socioeconomic features (see Table 10). This is a significant finding, which suggests that socioeconomic status was more effective in predicting 2021 cumulative cases than a country's median age, population density, and vaccination uptake, highlighting its unique importance as a nonclinical predictor of COVID-19 in the sample of countries.

**Table 8.** Feature Model 1 comparison (outliers included versus excluded).

Algorithm	Mean $R^2$ in the sample with outliers included (n=182)	Mean $R^2$ in the sample with outliers excluded (n=178)
Linear regression	0.693	0.689
Random forest	0.481	0.493
AdaBoost	0.454	0.494

**Table 9.** Feature Model 2 comparison (outliers included versus excluded).

Algorithm	Mean $R^2$ in the sample with outliers included (n=182)	Mean $R^2$ in the sample with outliers excluded (n=178)
Linear regression	0.763	0.754
Random forest	0.722	0.777
AdaBoost	0.679	0.733

**Table 10.** Feature Model 2 performance comparison of socioeconomic metrics versus other risk factors using linear regression.

Measure	Feature Model 2 with HDI <sup>a</sup> indicators	Feature Model 2 with population density, median age, and vaccination uptake
R <sup>2</sup>	0.763	0.661
MAE <sup>b</sup>	0.070	0.075
MSE <sup>c</sup>	0.011	0.016
RMSE <sup>d</sup>	0.107	0.128
Max error	0.265	0.312

<sup>a</sup>HDI: Human Development Index.

<sup>b</sup>MAE: mean absolute error.

<sup>c</sup>MSE: mean squared error.

<sup>d</sup>RMSE: root mean squared error.

## Significance of the Results

In order to put the machine learning results of this study into perspective, we compared the best performing algorithm (R<sup>2</sup>=0.763) with similar machine learning COVID-19 case predictions. Overall, it fits within the accepted range of COVID-19 predictive modeling studies in the systematic review mentioned earlier, which ranged from 0.64 to 1 [64]. Results from this study align with the findings from another study that attempted to predict COVID-19 cumulative cases in 3109 counties in the United States using a multilayer perceptron neural network. In this previous study, the socioeconomic indicator of median household income ranked fifth among 57 clinical and nonclinical predictor variables of COVID-19 prevalence [88]. Studies, such as this, portray the importance of socioeconomic indicators as determinants of COVID-19 prevalence rates, which further supports the use of HDI in this study to more accurately and precisely predict COVID-19 prevalence in 2021.

This research has a number of implications. First, it showcases the utility of combining statistical and machine learning approaches in pandemic research. Although statistical tests can determine correlations between variables, they cannot provide specific predictions of the target variable. Each method thus addresses a shortcoming of the other. Second, this study indicates that socioeconomic status is an important variable to consider in future epidemiological modeling, and reveals the complex social nature of the COVID-19 pandemic. Socioeconomic status was a better predictor of COVID-19 prevalence than median age, population density, and vaccination uptake. Third, the accuracy of these results in a multicountry sample is noteworthy. Owing to the data taken from 182 countries, this research suggests that socioeconomic status can be considered a “global risk factor” rather than a country-specific factor [38]. This will support evidence-based policy and interventions by decision makers. Fourth, the results indicate that although socioeconomic factors aid in the prediction of COVID-19, there could be other important factors that could further optimize prediction. Finally, the importance of historically reported COVID-19 case data cannot be

understated in attempting to predict future COVID-19 prevalence. The 2020 COVID-19 case data correlated strongly with 2021 COVID-19 case data and could be considered the most important machine learning feature.

## Limitations

As with all research studies, there are inherent limitations in this study. First, when analyzing COVID-19 cross-nationally, it must be noted that some countries have underreported their number of cases more than others for reasons, such as limited testing capacity [89]. Second, there are other socioeconomic factors that the HDI does not account for, including levels of financial inequality, social exclusion, or discrimination within countries [90]. These factors are worthy of inclusion in future research to assess their impact. Third, national COVID-19 prevalence rates give an overall measure of how severely a country is impacted, which is suitable for cross-country research, but they do not capture the full complexity of transmission patterns within each country. It is recommended that further research be conducted at the regional and municipal levels to assist pandemic forecasting. Lastly, it can be challenging to train reliable machine learning models using small data sets [59]. Cross-validation was used to address this limitation, as it maximizes the data set and minimizes the potential bias of a traditional partitioning approach.

## Conclusions

A better understanding of population-level predictors is of crucial importance to better understand and respond to public health crises caused by COVID-19 [91]. This study contributes to the growing corpus of COVID-19 predictive modeling research by showing that socioeconomic status is an important nonclinical risk factor. Using HDI and historical case rates, it was observed that 2021 cross-national COVID-19 cumulative cases could be predicted with a reasonable degree of accuracy. Although COVID-19 represents a long-term challenge for the global society, the data-driven approach of machine learning will continue to support decision makers in understanding the pandemic, formulating response strategies, and predicting future outcomes [92].

## Conflicts of Interest

None declared.

## References

1. The largest global rollout of vaccines in history just got one step closer. Gavi, The Vaccine Alliance. URL: <https://www.gavi.org/vaccineswork/largest-global-rollout-vaccines-history-just-got-one-step-closer> [accessed 2021-11-05]
2. Iyanda AE, Adeleke R, Lu Y, Osayomi T, Adaralegbe A, Lasode M, et al. A retrospective cross-national examination of COVID-19 outbreak in 175 countries: a multiscale geographically weighted regression analysis (January 11-June 28, 2020). *J Infect Public Health* 2020 Oct;13(10):1438-1445 [FREE Full text] [doi: [10.1016/j.jiph.2020.07.006](https://doi.org/10.1016/j.jiph.2020.07.006)] [Medline: [32773211](https://pubmed.ncbi.nlm.nih.gov/32773211/)]
3. Balmford B, Annan JD, Hargreaves JC, Altoè M, Bateman IJ. Cross-Country Comparisons of Covid-19: Policy, Politics and the Price of Life. *Environ Resour Econ (Dordr)* 2020;76(4):525-551 [FREE Full text] [doi: [10.1007/s10640-020-00466-5](https://doi.org/10.1007/s10640-020-00466-5)] [Medline: [32836862](https://pubmed.ncbi.nlm.nih.gov/32836862/)]
4. Földi M, Farkas N, Kiss S, Zádori N, Vánca S, Szakó L, KETLAK Study Group. Obesity is a risk factor for developing critical condition in COVID-19 patients: A systematic review and meta-analysis. *Obes Rev* 2020 Oct;21(10):e13095 [FREE Full text] [doi: [10.1111/obr.13095](https://doi.org/10.1111/obr.13095)] [Medline: [32686331](https://pubmed.ncbi.nlm.nih.gov/32686331/)]
5. Mahase E. Covid-19: Why are age and obesity risk factors for serious disease? *BMJ* 2020 Oct 26;371:m4130. [doi: [10.1136/bmj.m4130](https://doi.org/10.1136/bmj.m4130)] [Medline: [33106243](https://pubmed.ncbi.nlm.nih.gov/33106243/)]
6. Masood M, Aggarwal A, Reidpath DD. Effect of national culture on BMI: a multilevel analysis of 53 countries. *BMC Public Health* 2019 Sep 03;19(1):1212 [FREE Full text] [doi: [10.1186/s12889-019-7536-0](https://doi.org/10.1186/s12889-019-7536-0)] [Medline: [31481044](https://pubmed.ncbi.nlm.nih.gov/31481044/)]
7. Zhou Y, Chi J, Lv W, Wang Y. Obesity and diabetes as high-risk factors for severe coronavirus disease 2019 (Covid-19). *Diabetes Metab Res Rev* 2021 Feb;37(2):e3377 [FREE Full text] [doi: [10.1002/dmrr.3377](https://doi.org/10.1002/dmrr.3377)] [Medline: [32588943](https://pubmed.ncbi.nlm.nih.gov/32588943/)]
8. Lima-Martínez MM, Carrera Boada C, Madera-Silva MD, Marín W, Contreras M. COVID-19 and diabetes: A bidirectional relationship. *Clin Investig Arterioscler* 2021;33(3):151-157 [FREE Full text] [doi: [10.1016/j.arteri.2020.10.001](https://doi.org/10.1016/j.arteri.2020.10.001)] [Medline: [33303218](https://pubmed.ncbi.nlm.nih.gov/33303218/)]
9. Kashyap VK, Dhasmana A, Massey A, Kotnala S, Zafar N, Jaggi M, et al. Smoking and COVID-19: Adding Fuel to the Flame. *Int J Mol Sci* 2020 Sep 09;21(18):6581 [FREE Full text] [doi: [10.3390/ijms21186581](https://doi.org/10.3390/ijms21186581)] [Medline: [32916821](https://pubmed.ncbi.nlm.nih.gov/32916821/)]
10. Ibanez A, Sisodia GS. The role of culture on 2020 SARS-CoV-2 Country deaths: a pandemic management based on cultural dimensions. *GeoJournal* 2022;87(2):1175-1191 [FREE Full text] [doi: [10.1007/s10708-020-10306-0](https://doi.org/10.1007/s10708-020-10306-0)] [Medline: [33020679](https://pubmed.ncbi.nlm.nih.gov/33020679/)]
11. Yeung AY, Roewer-Despres F, Rosella L, Rudzicz F. Machine Learning-Based Prediction of Growth in Confirmed COVID-19 Infection Cases in 114 Countries Using Metrics of Nonpharmaceutical Interventions and Cultural Dimensions: Model Development and Validation. *J Med Internet Res* 2021 Apr 23;23(4):e26628 [FREE Full text] [doi: [10.2196/26628](https://doi.org/10.2196/26628)] [Medline: [33844636](https://pubmed.ncbi.nlm.nih.gov/33844636/)]
12. Cascini F, Pantovic A, Al-Ajlouni Y, Failla G, Ricciardi W. Attitudes, acceptance and hesitancy among the general population worldwide to receive the COVID-19 vaccines and their contributing factors: A systematic review. *EClinicalMedicine* 2021 Oct;40:101113 [FREE Full text] [doi: [10.1016/j.eclinm.2021.101113](https://doi.org/10.1016/j.eclinm.2021.101113)] [Medline: [34490416](https://pubmed.ncbi.nlm.nih.gov/34490416/)]
13. Troumbis AY. Testing the socioeconomic determinants of COVID-19 pandemic hypothesis with aggregated Human Development Index. *J Epidemiol Community Health* 2020 Dec 08;jech-2020-215986. [doi: [10.1136/jech-2020-215986](https://doi.org/10.1136/jech-2020-215986)] [Medline: [33293289](https://pubmed.ncbi.nlm.nih.gov/33293289/)]
14. Azza A, Sarhan A. Using Human Development Indices to Identify Indicators to Monitor the Corona Virus Pandemic. *JVAT* 2020 Apr 17;1(1):48-57. [doi: [10.14302/issn.2691-8862.jvat-20-3306](https://doi.org/10.14302/issn.2691-8862.jvat-20-3306)]
15. de Souza CDF, Machado MF, do Carmo RF. Human development, social vulnerability and COVID-19 in Brazil: a study of the social determinants of health. *Infect Dis Poverty* 2020 Aug 31;9(1):124 [FREE Full text] [doi: [10.1186/s40249-020-00743-x](https://doi.org/10.1186/s40249-020-00743-x)] [Medline: [32867851](https://pubmed.ncbi.nlm.nih.gov/32867851/)]
16. Stanton EA. The Human Development Index: A History. University of Massachusetts-Amherst. 2007. URL: [https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1101&context=peri\\_workingpapers](https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1101&context=peri_workingpapers) [accessed 2021-07-05]
17. Franco-Herrera D, González-Ocampo D, Restrepo-Montoya V, Gómez-Guevara JE, Alvear-Villacorte N, Rodríguez-Morales AJ. Relationship between malaria epidemiology and the human development index in Colombia and Latin America. *Infez Med* 2018 Sep 01;26(3):255-262 [FREE Full text] [Medline: [30246769](https://pubmed.ncbi.nlm.nih.gov/30246769/)]
18. Pervaiz R, Faisal F. Investigating the Nexus between Gynaecologic Cancer and Human Development Index. *Afr J Reprod Health* 2020 Mar;24(1):53-61. [doi: [10.29063/ajrh2020/v24i1.6](https://doi.org/10.29063/ajrh2020/v24i1.6)] [Medline: [32358937](https://pubmed.ncbi.nlm.nih.gov/32358937/)]
19. Khazaei Z, Goodarzi E, Borhaninejad V, Iranmanesh F, Mirshekarpour H, Mirzaei B, et al. The association between incidence and mortality of brain cancer and human development index (HDI): an ecological study. *BMC Public Health* 2020 Nov 12;20(1):1696 [FREE Full text] [doi: [10.1186/s12889-020-09838-4](https://doi.org/10.1186/s12889-020-09838-4)] [Medline: [33183267](https://pubmed.ncbi.nlm.nih.gov/33183267/)]
20. Zeng Z, Chen J, Xiao C, Chen W. A Global View on Prevalence of Hypertension and Human Development Index. *Ann Glob Health* 2020 Jun 29;86(1):67 [FREE Full text] [doi: [10.5334/aogh.2591](https://doi.org/10.5334/aogh.2591)] [Medline: [32676296](https://pubmed.ncbi.nlm.nih.gov/32676296/)]
21. Singu S, Acharya A, Challagundla K, Byrareddy SN. Impact of Social Determinants of Health on the Emerging COVID-19 Pandemic in the United States. *Front Public Health* 2020 Jul 21;8:406 [FREE Full text] [doi: [10.3389/fpubh.2020.00406](https://doi.org/10.3389/fpubh.2020.00406)] [Medline: [32793544](https://pubmed.ncbi.nlm.nih.gov/32793544/)]

22. Darin-Mattsson A, Fors S, Kåreholt I. Different indicators of socioeconomic status and their relative importance as determinants of health in old age. *Int J Equity Health* 2017 Sep 26;16(1):173 [FREE Full text] [doi: [10.1186/s12939-017-0670-3](https://doi.org/10.1186/s12939-017-0670-3)] [Medline: [28950875](https://pubmed.ncbi.nlm.nih.gov/28950875/)]
23. Hellmich SN. What is Socioeconomics? An Overview of Theories, Methods, and Themes in the Field. *Forum for Social Economics* 2015 Jan 15;46(1):3-25. [doi: [10.1080/07360932.2014.999696](https://doi.org/10.1080/07360932.2014.999696)]
24. Adler NE, Boyce T, Chesney MA, Cohen S, Folkman S, Kahn RL, et al. Socioeconomic status and health. The challenge of the gradient. *Am Psychol* 1994 Jan;49(1):15-24. [doi: [10.1037//0003-066x.49.1.15](https://doi.org/10.1037//0003-066x.49.1.15)] [Medline: [8122813](https://pubmed.ncbi.nlm.nih.gov/8122813/)]
25. Wang J, Geng L. Effects of Socioeconomic Status on Physical and Psychological Health: Lifestyle as a Mediator. *Int J Environ Res Public Health* 2019 Jan 20;16(2):281 [FREE Full text] [doi: [10.3390/ijerph16020281](https://doi.org/10.3390/ijerph16020281)] [Medline: [30669511](https://pubmed.ncbi.nlm.nih.gov/30669511/)]
26. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep* 2014;129 Suppl 2:19-31 [FREE Full text] [doi: [10.1177/00333549141291S206](https://doi.org/10.1177/00333549141291S206)] [Medline: [24385661](https://pubmed.ncbi.nlm.nih.gov/24385661/)]
27. Walters S, Suhrcke M. Socioeconomic inequalities in health and health care access in central and eastern Europe and the CIS: a review of the recent literature. World Health Organization. 2005. URL: <https://apps.who.int/iris/handle/10665/350352> [accessed 2022-03-01]
28. Adler NE, Ostrove JM. Socioeconomic status and health: what we know and what we don't. *Ann N Y Acad Sci* 1999;896:3-15. [doi: [10.1111/j.1749-6632.1999.tb08101.x](https://doi.org/10.1111/j.1749-6632.1999.tb08101.x)] [Medline: [10681884](https://pubmed.ncbi.nlm.nih.gov/10681884/)]
29. Hakeberg M, Wide Boman U. Self-reported oral and general health in relation to socioeconomic position. *BMC Public Health* 2017 Jul 26;18(1):63 [FREE Full text] [doi: [10.1186/s12889-017-4609-9](https://doi.org/10.1186/s12889-017-4609-9)] [Medline: [28747180](https://pubmed.ncbi.nlm.nih.gov/28747180/)]
30. Hawkins RB, Charles EJ, Mehaffey JH. Socio-economic status and COVID-19-related cases and fatalities. *Public Health* 2020 Dec;189:129-134 [FREE Full text] [doi: [10.1016/j.puhe.2020.09.016](https://doi.org/10.1016/j.puhe.2020.09.016)] [Medline: [33227595](https://pubmed.ncbi.nlm.nih.gov/33227595/)]
31. Patel JA, Nielsen FBH, Badiani AA, Assi S, Unadkat VA, Patel B, et al. Poverty, inequality and COVID-19: the forgotten vulnerable. *Public Health* 2020 Jun;183:110-111 [FREE Full text] [doi: [10.1016/j.puhe.2020.05.006](https://doi.org/10.1016/j.puhe.2020.05.006)] [Medline: [32502699](https://pubmed.ncbi.nlm.nih.gov/32502699/)]
32. Human Development Report 2020. United Nations Development Programme. 2020. URL: <https://hdr.undp.org/system/files/documents/hdr2020pdf.pdf> [accessed 2021-09-08]
33. Hu Q, Zhang Q, Chen W, Bai X, Liang T. Human development index is associated with mortality-to-incidence ratios of gastrointestinal cancers. *World J Gastroenterol* 2013 Aug 28;19(32):5261-5270 [FREE Full text] [doi: [10.3748/wjg.v19.i32.5261](https://doi.org/10.3748/wjg.v19.i32.5261)] [Medline: [23983428](https://pubmed.ncbi.nlm.nih.gov/23983428/)]
34. Soheylizad M, Khazaei S, Jenabi E, Delpisheh A, Veisani Y. The Relationship Between Human Development Index and Its Components with Thyroid Cancer Incidence and Mortality: Using the Decomposition Approach. *Int J Endocrinol Metab* 2018 Oct;16(4):e65078 [FREE Full text] [doi: [10.5812/ijem.65078](https://doi.org/10.5812/ijem.65078)] [Medline: [30464773](https://pubmed.ncbi.nlm.nih.gov/30464773/)]
35. Javanmard E, Niyiyati M, Ghasemi E, Mirjalali H, Asadzadeh Aghdaei H, Zali MR. Impacts of human development index and climate conditions on prevalence of Blastocystis: A systematic review and meta-analysis. *Acta Trop* 2018 Sep;185:193-203. [doi: [10.1016/j.actatropica.2018.05.014](https://doi.org/10.1016/j.actatropica.2018.05.014)] [Medline: [29802845](https://pubmed.ncbi.nlm.nih.gov/29802845/)]
36. Pereira FA, de Mendonça IA, Werneck RI, Moysés ST, Gabardo MC, Moysés SJ. Human Development Index, Ratio of Dentists and Inhabitants, and the Decayed, Missing or Filled Teeth Index in Large Cities. *J Contemp Dent Pract* 2018 Nov 01;19(11):1363-1369. [Medline: [30602642](https://pubmed.ncbi.nlm.nih.gov/30602642/)]
37. Shaikh A, Peprah E, Mohamed RH, Asghar A, Andharia NV, Lajot NA, et al. COVID-19 and mental health: a multi-country study—the effects of lockdown on the mental health of young adults. *Middle East Curr Psychiatry* 2021 Aug 09;28(1):1-10. [doi: [10.1186/s43045-021-00116-6](https://doi.org/10.1186/s43045-021-00116-6)]
38. O'Sullivan R, Burns A, Leavey G, Leroi I, Burholt V, Lubben J, et al. Impact of the COVID-19 Pandemic on Loneliness and Social Isolation: A Multi-Country Study. *Int J Environ Res Public Health* 2021 Sep 23;18(19):9982 [FREE Full text] [doi: [10.3390/ijerph18199982](https://doi.org/10.3390/ijerph18199982)] [Medline: [34639283](https://pubmed.ncbi.nlm.nih.gov/34639283/)]
39. Hawlader MDH, Rahman ML, Nazir A, Ara T, Haque MMA, Saha S, et al. COVID-19 vaccine acceptance in South Asia: a multi-country study. *Int J Infect Dis* 2022 Jan;114:1-10 [FREE Full text] [doi: [10.1016/j.ijid.2021.09.056](https://doi.org/10.1016/j.ijid.2021.09.056)] [Medline: [34597765](https://pubmed.ncbi.nlm.nih.gov/34597765/)]
40. Jang SY, Hussain-Alkhateeb L, Rivera Ramirez T, Al-Aghbari AA, Chackalackal DJ, Cardenas-Sanchez R, et al. Factors shaping the COVID-19 epidemic curve: a multi-country analysis. *BMC Infect Dis* 2021 Oct 02;21(1):1032 [FREE Full text] [doi: [10.1186/s12879-021-06714-3](https://doi.org/10.1186/s12879-021-06714-3)] [Medline: [34600485](https://pubmed.ncbi.nlm.nih.gov/34600485/)]
41. Chen Y, Lu P, Chang C, Liu T. A Time-Dependent SIR Model for COVID-19 With Undetectable Infected Persons. *IEEE Trans. Netw. Sci. Eng* 2020 Oct 1;7(4):3279-3294. [doi: [10.1109/tNSE.2020.3024723](https://doi.org/10.1109/tNSE.2020.3024723)]
42. Calafiore G, Novara C, Possieri C. A Modified SIR Model for the COVID-19 Contagion in Italy. 2020 Presented at: 59th IEEE Conference on Decision and Control (CDC); December 14-18, 2020; Jeju, Korea (South). [doi: [10.1109/cdc42340.2020.9304142](https://doi.org/10.1109/cdc42340.2020.9304142)]
43. Law KB, Peariasamy KM, Ibrahim H, Abdullah NH. Modelling infectious diseases with herd immunity in a randomly mixed population. Research Square. URL: <https://www.researchsquare.com/article/rs-289776/v5> [accessed 2022-09-03]
44. Kadkhoda K. Herd Immunity to COVID-19. *Am J Clin Pathol* 2021 Mar 15;155(4):471-472 [FREE Full text] [doi: [10.1093/ajcp/aqaa272](https://doi.org/10.1093/ajcp/aqaa272)] [Medline: [33399182](https://pubmed.ncbi.nlm.nih.gov/33399182/)]
45. Wiemken TL, Kelley RR. Machine Learning in Epidemiology and Health Outcomes Research. *Annu Rev Public Health* 2020 Apr 02;41:21-36. [doi: [10.1146/annurev-publhealth-040119-094437](https://doi.org/10.1146/annurev-publhealth-040119-094437)] [Medline: [31577910](https://pubmed.ncbi.nlm.nih.gov/31577910/)]

46. Anderson AB, Grazal CF, Balazs GC, Potter BK, Dickens JF, Forsberg JA. Can Predictive Modeling Tools Identify Patients at High Risk of Prolonged Opioid Use After ACL Reconstruction? *Clin Orthop Relat Res* 2020 Jul;478(7):1618 [FREE Full text] [doi: [10.1097/CORR.0000000000001251](https://doi.org/10.1097/CORR.0000000000001251)] [Medline: [32282466](https://pubmed.ncbi.nlm.nih.gov/32282466/)]
47. Kang AR, Lee J, Jung W, Lee M, Park SY, Woo J, et al. Development of a prediction model for hypotension after induction of anesthesia using machine learning. *PLoS One* 2020;15(4):e0231172 [FREE Full text] [doi: [10.1371/journal.pone.0231172](https://doi.org/10.1371/journal.pone.0231172)] [Medline: [32298292](https://pubmed.ncbi.nlm.nih.gov/32298292/)]
48. Shickel B, Loftus TJ, Adhikari L, Ozrazgat-Baslanti T, Bihorac A, Rashidi P. DeepSOFA: A Continuous Acuity Score for Critically Ill Patients using Clinically Interpretable Deep Learning. *Sci Rep* 2019 Feb 12;9(1):1879 [FREE Full text] [doi: [10.1038/s41598-019-38491-0](https://doi.org/10.1038/s41598-019-38491-0)] [Medline: [30755689](https://pubmed.ncbi.nlm.nih.gov/30755689/)]
49. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is Machine Learning? A Primer for the Epidemiologist. *Am J Epidemiol* 2019 Dec 31;188(12):2222-2239. [doi: [10.1093/aje/kwz189](https://doi.org/10.1093/aje/kwz189)] [Medline: [31509183](https://pubmed.ncbi.nlm.nih.gov/31509183/)]
50. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature* 2013 Apr 25;496(7446):504-507 [FREE Full text] [doi: [10.1038/nature12060](https://doi.org/10.1038/nature12060)] [Medline: [23563266](https://pubmed.ncbi.nlm.nih.gov/23563266/)]
51. Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, et al. Mapping the zoonotic niche of Ebola virus disease in Africa. *Elife* 2014 Sep 08;3:e04395 [FREE Full text] [doi: [10.7554/eLife.04395](https://doi.org/10.7554/eLife.04395)] [Medline: [25201877](https://pubmed.ncbi.nlm.nih.gov/25201877/)]
52. Messina JP, Pigott DM, Golding N, Duda KA, Brownstein JS, Weiss DJ, et al. The global distribution of Crimean-Congo hemorrhagic fever. *Trans R Soc Trop Med Hyg* 2015 Aug;109(8):503-513 [FREE Full text] [doi: [10.1093/trstmh/trv050](https://doi.org/10.1093/trstmh/trv050)] [Medline: [26142451](https://pubmed.ncbi.nlm.nih.gov/26142451/)]
53. Messina JP, Kraemer MU, Brady OJ, Pigott DM, Shearer FM, Weiss DJ, et al. Mapping global environmental suitability for Zika virus. *Elife* 2016 Apr 19;5:e15272 [FREE Full text] [doi: [10.7554/eLife.15272](https://doi.org/10.7554/eLife.15272)] [Medline: [27090089](https://pubmed.ncbi.nlm.nih.gov/27090089/)]
54. Shaman J, Karspeck A. Forecasting seasonal outbreaks of influenza. *Proc Natl Acad Sci U S A* 2012 Dec 11;109(50):20425-20430 [FREE Full text] [doi: [10.1073/pnas.1208772109](https://doi.org/10.1073/pnas.1208772109)] [Medline: [23184969](https://pubmed.ncbi.nlm.nih.gov/23184969/)]
55. Wang L, Yang G, Jia L, Li Z, Xie J, Li P, et al. Epidemiological features and trends of Ebola virus disease in West Africa. *Int J Infect Dis* 2015 Sep;38:52-53 [FREE Full text] [doi: [10.1016/j.ijid.2015.07.017](https://doi.org/10.1016/j.ijid.2015.07.017)] [Medline: [26216765](https://pubmed.ncbi.nlm.nih.gov/26216765/)]
56. DeFelice NB, Little E, Campbell SR, Shaman J. Ensemble forecast of human West Nile virus cases and mosquito infection rates. *Nat Commun* 2017 Feb 24;8:14592 [FREE Full text] [doi: [10.1038/ncomms14592](https://doi.org/10.1038/ncomms14592)] [Medline: [28233783](https://pubmed.ncbi.nlm.nih.gov/28233783/)]
57. Reis J, Shaman J. Retrospective Parameter Estimation and Forecast of Respiratory Syncytial Virus in the United States. *PLoS Comput Biol* 2016 Oct;12(10):e1005133 [FREE Full text] [doi: [10.1371/journal.pcbi.1005133](https://doi.org/10.1371/journal.pcbi.1005133)] [Medline: [27716828](https://pubmed.ncbi.nlm.nih.gov/27716828/)]
58. Painuli D, Mishra D, Bhardwaj S, Aggarwal M. Forecast and prediction of COVID-19 using machine learning. In: Kose U, Gupta D, de Albuquerque VHC, Khanna A, editors. *Data Science for COVID-19*. Cambridge, MA: Academic Press; 2021:381-397.
59. Ahmad A, Garhwal S, Ray SK, Kumar G, Malebary SJ, Barukab OM. The Number of Confirmed Cases of Covid-19 by using Machine Learning: Methods and Challenges. *Arch Comput Methods Eng* 2021 Aug 04;28(4):2645-2653 [FREE Full text] [doi: [10.1007/s11831-020-09472-8](https://doi.org/10.1007/s11831-020-09472-8)] [Medline: [32837183](https://pubmed.ncbi.nlm.nih.gov/32837183/)]
60. Assaf D, Gutman Y, Neuman Y, Segal G, Amit S, Gefen-Halevi S, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern Emerg Med* 2020 Nov;15(8):1435-1443 [FREE Full text] [doi: [10.1007/s11739-020-02475-0](https://doi.org/10.1007/s11739-020-02475-0)] [Medline: [32812204](https://pubmed.ncbi.nlm.nih.gov/32812204/)]
61. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, Northwell COVID-19 Research Consortium. A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation. *J Med Internet Res* 2021 Feb 10;23(2):e24246 [FREE Full text] [doi: [10.2196/24246](https://doi.org/10.2196/24246)] [Medline: [33476281](https://pubmed.ncbi.nlm.nih.gov/33476281/)]
62. Gao Y, Cai G, Fang W, Li H, Wang S, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat Commun* 2020 Oct 06;11(1):5033 [FREE Full text] [doi: [10.1038/s41467-020-18684-2](https://doi.org/10.1038/s41467-020-18684-2)] [Medline: [33024092](https://pubmed.ncbi.nlm.nih.gov/33024092/)]
63. Banoei MM, Dinparastisaleh R, Zadeh AV, Mirsaedi M. Machine-learning-based COVID-19 mortality prediction model and identification of patients at low and high risk of dying. *Crit Care* 2021 Sep 08;25(1):328 [FREE Full text] [doi: [10.1186/s13054-021-03749-5](https://doi.org/10.1186/s13054-021-03749-5)] [Medline: [34496940](https://pubmed.ncbi.nlm.nih.gov/34496940/)]
64. Ghafouri-Fard S, Mohammad-Rahimi H, Motie P, Minabi MAS, Taheri M, Nateghinia S. Application of machine learning in the prediction of COVID-19 daily new cases: A scoping review. *Heliyon* 2021 Oct;7(10):e08143 [FREE Full text] [doi: [10.1016/j.heliyon.2021.e08143](https://doi.org/10.1016/j.heliyon.2021.e08143)] [Medline: [34660935](https://pubmed.ncbi.nlm.nih.gov/34660935/)]
65. Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, et al. Coronavirus Pandemic (COVID-19). *Our World in Data*. URL: <https://ourworldindata.org/coronavirus> [accessed 2021-09-01]
66. Pollet TV, Stulp G, Henzi SP, Barrett L. Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data. *Am J Primatol* 2015 Jul;77(7):727-740. [doi: [10.1002/ajp.22405](https://doi.org/10.1002/ajp.22405)] [Medline: [25810242](https://pubmed.ncbi.nlm.nih.gov/25810242/)]
67. Data downloads. United Nations Development Programme. URL: <https://hdr.undp.org/data-center/documentation-and-downloads> [accessed 2022-03-10]
68. Ahsan MM, Mahmud MAP, Saha PK, Gupta KD, Siddique Z. Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies* 2021 Jul 24;9(3):52. [doi: [10.3390/technologies9030052](https://doi.org/10.3390/technologies9030052)]

69. Patel K, Drucker SM, Fogarty J, Kapoor A, Tan DS. Using Multiple Models to Understand Data. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. 2011 Presented at: Twenty-Second International Joint Conference on Artificial Intelligence; July 16-22, 2011; Barcelona, Catalonia, Spain. [doi: [10.5591/978-1-57735-516-8/IJCAI11-289](https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-289)]
70. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak* 2019 Dec 21;19(1):281 [FREE Full text] [doi: [10.1186/s12911-019-1004-8](https://doi.org/10.1186/s12911-019-1004-8)] [Medline: [31864346](https://pubmed.ncbi.nlm.nih.gov/31864346/)]
71. Xu Y, Goodacre R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *J Anal Test* 2018;2(3):249-262 [FREE Full text] [doi: [10.1007/s41664-018-0068-2](https://doi.org/10.1007/s41664-018-0068-2)] [Medline: [30842888](https://pubmed.ncbi.nlm.nih.gov/30842888/)]
72. Maulud D, Abdulazeez AM. A Review on Linear Regression Comprehensive in Machine Learning. *JASTT* 2020 Dec 31;1(4):140-147. [doi: [10.38094/jastt1457](https://doi.org/10.38094/jastt1457)]
73. Biau G. Analysis of a Random Forests Model. *Journal of Machine Learning Research* 2012;13:1063-1095 [FREE Full text]
74. Belle V, Papantonis I. Principles and Practice of Explainable Machine Learning. *Front Big Data* 2021 Jul 1;4:688969 [FREE Full text] [doi: [10.3389/fdata.2021.688969](https://doi.org/10.3389/fdata.2021.688969)] [Medline: [34278297](https://pubmed.ncbi.nlm.nih.gov/34278297/)]
75. Wyner AJ, Olson M, Bleich J, Mease D. Explaining the Success of AdaBoost and Random Forests as Interpolating Classifiers. *Journal of Machine Learning Research* 2017;18:1-33 [FREE Full text]
76. Marcot BG, Hanea AM. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Comput Stat* 2020 Jun 13;36(3):2009-2031. [doi: [10.1007/s00180-020-00999-9](https://doi.org/10.1007/s00180-020-00999-9)]
77. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *IJCAI'95: Proceedings of the 14th International Joint Conference on Artificial Intelligence*. 1995 Presented at: 14th International Joint Conference on Artificial Intelligence; August 20-25, 1995; Montreal, Quebec, Canada p. 1137-1143. [doi: [10.5555/1643031.1643047](https://doi.org/10.5555/1643031.1643047)]
78. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci* 2021;7:e623 [FREE Full text] [doi: [10.7717/peerj-cs.623](https://doi.org/10.7717/peerj-cs.623)] [Medline: [34307865](https://pubmed.ncbi.nlm.nih.gov/34307865/)]
79. Bayati M. Why Is COVID-19 More Concentrated in Countries with High Economic Status? *Iran J Public Health* 2021 Sep;50(9):1926-1929 [FREE Full text] [doi: [10.18502/ijph.v50i9.7081](https://doi.org/10.18502/ijph.v50i9.7081)] [Medline: [34722396](https://pubmed.ncbi.nlm.nih.gov/34722396/)]
80. Auret L, Aldrich C. Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals Engineering* 2012 Aug;35:27-42. [doi: [10.1016/j.mineng.2012.05.008](https://doi.org/10.1016/j.mineng.2012.05.008)]
81. Lo A, Chernoff H, Zheng T, Lo S. Why significant variables aren't automatically good predictors. *Proc Natl Acad Sci U S A* 2015 Nov 10;112(45):13892-13897 [FREE Full text] [doi: [10.1073/pnas.1518285112](https://doi.org/10.1073/pnas.1518285112)] [Medline: [26504198](https://pubmed.ncbi.nlm.nih.gov/26504198/)]
82. Breiman L. Random Forests. *Machine Learning* 2001;45:5-32. [doi: [10.1023/A1010933404324](https://doi.org/10.1023/A1010933404324)]
83. Wong DWS, Li Y. Spreading of COVID-19: Density matters. *PLoS One* 2020;15(12):e0242398 [FREE Full text] [doi: [10.1371/journal.pone.0242398](https://doi.org/10.1371/journal.pone.0242398)] [Medline: [33362283](https://pubmed.ncbi.nlm.nih.gov/33362283/)]
84. Davies NG, Klepac P, Liu Y, Prem K, Jit M, CMMID COVID-19 working group, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med* 2020 Aug;26(8):1205-1211. [doi: [10.1038/s41591-020-0962-9](https://doi.org/10.1038/s41591-020-0962-9)] [Medline: [32546824](https://pubmed.ncbi.nlm.nih.gov/32546824/)]
85. Wilder-Smith A. What is the vaccine effect on reducing transmission in the context of the SARS-CoV-2 delta variant? *The Lancet Infectious Diseases* 2022 Feb;22(2):152-153. [doi: [10.1016/s1473-3099\(21\)00690-3](https://doi.org/10.1016/s1473-3099(21)00690-3)]
86. Countries in the world by population. *Worldometers*. URL: <https://www.worldometers.info/world-population/population-by-country/> [accessed 2022-04-01]
87. *Factbook-Countries*. *IndexMundi*. URL: <https://www.indexmundi.com/factbook/countries> [accessed 2022-04-02]
88. Mollalo A, Rivera KM, Vahedi B. Artificial Neural Network Modeling of Novel Coronavirus (COVID-19) Incidence Rates across the Continental United States. *Int J Environ Res Public Health* 2020 Jun 12;17(12):4204 [FREE Full text] [doi: [10.3390/ijerph17124204](https://doi.org/10.3390/ijerph17124204)] [Medline: [32545581](https://pubmed.ncbi.nlm.nih.gov/32545581/)]
89. Lau H, Khosrawipour T, Kocbach P, Ichii H, Bania J, Khosrawipour V. Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. *Pulmonology* 2021;27(2):110-115 [FREE Full text] [doi: [10.1016/j.pulmoe.2020.05.015](https://doi.org/10.1016/j.pulmoe.2020.05.015)] [Medline: [32540223](https://pubmed.ncbi.nlm.nih.gov/32540223/)]
90. Bilbao-Ubillos J. The Limits of Human Development Index: The Complementary Role of Economic and Social Cohesion, Development Strategies and Sustainability. *Sust. Dev* 2011 May 19;21(6):400-412. [doi: [10.1002/sd.525](https://doi.org/10.1002/sd.525)]
91. Erman A, Medeiros M. Exploring the Effect of Collective Cultural Attributes on Covid-19-Related Public Health Outcomes. *Front Psychol* 2021;12:627669 [FREE Full text] [doi: [10.3389/fpsyg.2021.627669](https://doi.org/10.3389/fpsyg.2021.627669)] [Medline: [33833717](https://pubmed.ncbi.nlm.nih.gov/33833717/)]
92. Polonsky JA, Baidjoe A, Kamvar ZN, Cori A, Durski K, Edmunds WJ, et al. Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philos Trans R Soc Lond B Biol Sci* 2019 Jul 08;374(1776):20180276 [FREE Full text] [doi: [10.1098/rstb.2018.0276](https://doi.org/10.1098/rstb.2018.0276)] [Medline: [31104603](https://pubmed.ncbi.nlm.nih.gov/31104603/)]

## Abbreviations

**GNI:** gross national income



**HDI:** Human Development Index  
**MAE:** mean absolute error  
**MSE:** mean squared error  
**RMSE:** root mean squared error  
**SIR:** susceptible, infected, recovered  
**UNDP:** United Nations Development Programme

*Edited by A Mavragani; submitted 22.11.21; peer-reviewed by P Wang, S Rostam Niakan Kalhori, M Pradhan; comments to author 24.02.22; revised version received 12.04.22; accepted 27.04.22; published 27.09.22.*

*Please cite as:*

*Winston L, McCann M, Onofrei G*

*Exploring Socioeconomic Status as a Global Determinant of COVID-19 Prevalence, Using Exploratory Data Analytic and Supervised Machine Learning Techniques: Algorithm Development and Validation Study*

*JMIR Form Res 2022;6(9):e35114*

*URL: <https://formative.jmir.org/2022/9/e35114>*

*doi: [10.2196/35114](https://doi.org/10.2196/35114)*

*PMID: [36001798](https://pubmed.ncbi.nlm.nih.gov/36001798/)*

©Luke Winston, Michael McCann, George Onofrei. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 27.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques

Muhammad Khubayeeb Kabir<sup>1\*</sup>, BSc; Maisha Islam<sup>1\*</sup>, BSc; Anika Nahian Binte Kabir<sup>1\*</sup>, BSc; Adiba Haque<sup>1\*</sup>, BSc; Md Khalilur Rhaman<sup>2</sup>, BSc, MSc, PhD

<sup>1</sup>Department of Computer Science, Brac University, Dhaka, Bangladesh

<sup>2</sup>Department of Computer Science and Engineering, Brac University, Dhaka, Bangladesh

\*these authors contributed equally

**Corresponding Author:**

Muhammad Khubayeeb Kabir, BSc

Department of Computer Science

Brac University

66 Mohakhali

Dhaka, 1212

Bangladesh

Phone: 880 1708812609

Email: [muhammad.khubayeeb.kabir@g.bracu.ac.bd](mailto:muhammad.khubayeeb.kabir@g.bracu.ac.bd)

## Abstract

**Background:** There are a myriad of language cues that indicate depression in written texts, and natural language processing (NLP) researchers have proven the ability of machine learning and deep learning approaches to detect these cues. However, to date, these approaches bridging NLP and the domain of mental health for Bengali literature are not comprehensive. The Bengali-speaking population can express emotions in their native language in greater detail.

**Objective:** Our goal is to detect the severity of depression using Bengali texts by generating a novel Bengali corpus of depressive posts. We collaborated with mental health experts to generate a clinically sound labeling scheme and an annotated corpus to train machine learning and deep learning models.

**Methods:** We conducted a study using Bengali text-based data from blogs and open source platforms. We constructed a procedure for annotated corpus generation and extraction of textual information from Bengali literature for predictive analysis. We developed our own structured data set and designed a clinically sound labeling scheme with the help of mental health professionals, adhering to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)* during the process. We used 5 machine learning models for detecting the severity of depression: kernel support vector machine (SVM), random forest, logistic regression K-nearest neighbor (KNN), and complement naive Bayes (NB). For the deep learning approach, we used long short-term memory (LSTM) units and gated recurrent units (GRUs) coupled with convolutional blocks or self-attention layers. Finally, we aimed for enhanced outcomes by using state-of-the-art pretrained language models.

**Results:** The independent recurrent neural network (RNN) models yielded the highest accuracies and weighted F1 scores. GRUs, in particular, produced 81% accuracy. The hybrid architectures could not surpass the RNNs in terms of performance. Kernel SVM with term frequency-inverse document frequency (TF-IDF) embeddings generated 78% accuracy on test data. We used validation and training loss curves to observe and report the performance of our architectures. Overall, the number of available data remained the limitation of our experiment.

**Conclusions:** The findings from our experimental setup indicate that machine learning and deep learning models are fairly capable of assessing the severity of mental health issues from texts. For the future, we suggest more research endeavors to increase the volume of Bengali text data, in particular, so that modern architectures reach improved generalization capability.



(JMIR Form Res 2022;6(9):e36118) doi:[10.2196/36118](https://doi.org/10.2196/36118)

## KEYWORDS

mental health forums; natural language processing; severity; major depressive disorder; deep learning; machine learning; multiclass text classification

## Introduction

Major depressive disorder (MDD) is a mental condition characterized by chronic low mood or lack of interest, with a slew of other concerning symptoms over a 2-week period. Depression afflicts an estimated 1 in 15 adults and young adults in a year [1] and is the leading cause of suicide, which is the second-leading cause of mortality worldwide [2,3]. The problem of depression became more pronounced during the COVID-19 lockdown, exacerbating the mental health issues experienced by individuals. It is also a reasonably complex disease to treat because people who suffer from depression are often hesitant to report such symptoms as mental illness remains highly stigmatized in many societies [4]. There exists an abundance of user data or cues related to mental health that can be used by experts to solve such chronic issues.

Data mining in mental health is an advancing field of study that involves the use of machine learning, deep learning, linguistic, and statistical techniques to find patterns in data. Researchers are faced with a range of options with regard to corpus generation from depressive texts: standard academic documents or texts about depression, hashtags, and user posts. Twitter, Facebook, Reddit, and blogs are platforms containing collections of naturally occurring texts. Textual data of internet users depicting symptoms, experiences, thoughts, and conversations about mental health are dispersed across various platforms. In the recent years, it has been observed that individuals, wary of the stigma, prefer to seek clinical help anonymously through writing on platforms, such as Twitter, Reddit, TalkSpace, BeyondMeds, and other social blogs that can connect users with health professionals, counselors, and other users who share similar experiences. eRisk is a specialized platform focused on analyzing early risks using users' texts; Losada et al [5] discussed ways of early detection of depression on the internet. Individuals resort to the option of creating anonymous posts soliciting advice for their conditions in special groups. These posts are often grouped under tags, topics, or even hashtags, such as  (#psychology\_and\_mind) and  (#help\_post), or in other cases, the social media groups may be dedicated solely to a particular mental health topic.

In psychiatry, symptoms of subjects are generally classified using predefined scales. The Hamilton Depression Rating Scale [6] is an instrument used in scaling depressive disorders. Over the years, various rating scales have emerged that assess symptoms to produce a diagnosis or score. Internet users present various symptoms through texts and often on niche platforms. Extracting these textual data pertaining to mental health and structuring them meaningfully are challenging tasks. A corpus must be generated by domain experts for it to hold validity in the field of psychiatric assessment. For training deep learning architectures, accurately annotated corpora are indispensable.

Ellendorf et al [7] proposed the PsyMine corpus, which was generated by domain experts and their agreement scores were

presented. Alonso et al [8] presented a comprehensive review of data mining techniques in the mental health domain. In doing so, the authors covered depression, bipolar disorder, and schizophrenia. Reddit is a standard platform that compartmentalizes mental health posts into subreddits, such as SuicideWatch, bipolar disorder, and anxiety. The Reddit Self-Reported Diagnosis data set is a corpus comprising texts of 9000 reddit users. The corpus was generated with systematic user selection, and the annotation process was crowdsourced [9]. MacAveney et al [10] proposed RSDD-Time, which is a temporal corpus of self-proclaimed diagnosis statements. For each of the statements, the time of diagnosis and whether a condition is present were labeled. Additionally, the authors explored several classification approaches.

The fundamental objectives of this research are to generate an equivalent corpus for the Bengali language and to analyze the data set to detect depression and its severity in individuals. In our research, we integrated natural language processing (NLP) with machine learning and deep learning approaches. Prior work in the field of NLP demonstrated that machine learning and deep learning algorithms are capable of detecting depression-related cues in language [11,12]; however, to date, these efforts have focused on classifying the categories of mental illness rather than their degree. In our research, we adopted unique approaches to detecting and estimating the severity of depression, enabling us to identify those with depression on social media and safeguard others from viewing potentially triggering written content. Furthermore, the prior literature included the identification or categorization of mental diseases from texts in English, German, Russian, and other languages. Bengali is the fourth-most widely spoken language. Hence, we consolidated a process for textual information extraction from Bengali texts and performed lexical and predictive analysis for the purpose of detection of severity of depression.

Several studies in the field of multiclass emotion identification have been conducted using lexicon-based, machine learning, and deep learning approaches. A proposed approach by Mageed and Ungar [13] used gated recurrent neural networks (RNNs) to classify tweets into 24 emotion categories. They yielded over 80%  $F_1$  scores for some categories. Yang et al [14] and Ive et al [15] used a hierarchical architecture with a series of bidirectional encoders to classify different classes of mental health topics. Over the years, focus has shifted toward detection of depression of social media users. Cohan et al [16] created a self-reported depression data set to analyze the language usage of depressed users. They constructed a seed list of keywords assigned to the classes in their data set and applied a Linguistic Inquiry and Word Count (LIWC) approach to compare language usage between a user with and without depression. The experiment also involved the categorization of user posts using logistic regression, extreme gradient boosting (XGBoost), and convolutional neural networks (CNNs) into classes of mental disorders, namely attention deficit hyperactivity disorder (ADHD), bipolar disorder, posttraumatic stress disorder (PTSD),

and obsessive-compulsive disorder (OCD). Mustafa et al [17] proposed a novel approach to categorizing depressive texts in English using the LIWC text analysis technique. The posts collected belonged to specific Twitter hashtags, and the authors annotated depressive posts into 3 levels of severity: high, medium, and low. Words associated with mental illnesses were assigned weights, and a support vector machine (SVM) classifier [18], random forest, and 1D CNNs were used in the work. The usage of machine learning and deep learning techniques to scale the level of a condition or situation was proven quite feasible in recent research. For instance, Al-Garadi et al [19] used transformer models and CNNs to classify mentions of drug usage in English into 4 levels with the help of toxicologists. Identification of depressive texts in the Bengali language has been explored over the recent years through binary text classification techniques. Uddin et al [20] used RNNs to distinguish depressive and nondepressive texts. In the process, they fine-tuned the number of LSTM layers used. Moreover, Khan et al [21] collected Bengali text data from social media and blog posts to assemble a comprehensive Bengali data set containing expressions of positive and negative emotions.

We implemented several baseline models, such as kernel SVMs, complement naive Bayes (NB), logistic regression, random forest, and KNNs. Next, for a deep learning approach, we experimented with convolutional blocks and layers combined with RNNs. Among all the samples classified by bidirectional gated recurrent units (BiGRUs), 81% were correctly identified labels. Bidirectional long short-term memory (BiLSTM) classified 77% of the posts into the correct severity scale. We also reported the results of metrics, such as recall and  $F_1$  scores. In addition, we further explored bidirectional encoder representations from transformers (BERT) models using a pretrained monolingual XLM-RoBERTa language model [22] and expanded on the findings from these techniques.







Various approaches have been followed to identify or categorize depression using English texts, such as multilabel classification of mental disorders and identification of the severity of depression. However, in the context of the Bengali language, only binary classification approaches have been considered and are based solely on the polarity of emotions (ie, happy or sad). This prompted us to conduct research on the identification of hierarchical stages of depressive traits from the literature. For our research, we collected Bengali text data from similar microblogs or accessible social media groups. Generally, social blogs, forums, or groups have mechanisms that classify posts under specific topics pertaining to mental health, and often, these tags are all-encompassing or broad. The social blogs pertaining to mental health would benefit from a hierarchical classification mechanism, where user posts are addressed by professionals or experienced individuals based on urgency. Our technique is unique in the sense that it categorizes a spectrum of negative emotions from a novel Bengali language corpus of self-declared depressive symptoms and emotions. The individual texts were collected from various sources using a web-scraping Application Programming Interface (API) and categorized into 4 levels of severity by experts. Our code has been made publicly available [23].

## Methods

### Study Design

Our approach was twofold. First, we constructed a novel corpus of Bengali texts, consisting of posts exhibiting emotions or symptoms associated with mental illnesses. We studied recommended manuals for the assessment and diagnosis of medical depression to devise a scheme for data annotation. Second, we trained machine learning and deep learning models to classify the Bengali posts according to our scheme.

### Data Set



The Bengali posts were collected from social media platforms and blogs. We used Selenium, a Python web-scraping API, to collect data that originally consisted of code-mixed texts as well as pure Bengali and English texts. Some of the microblogs and social media groups that were relevant to our research included Monojogimon and  (schizophrenia) among many others. We automated our program to obtain posts under specific tags or topics. This assisted us in excluding posts that belonged to completely different topics. For blogs and microblogs, the filtering process included separating streams into different topics, such as  (depression),  (MDD),  (despair), and others. Among other resources, we collected user posts from Facebook groups, such as  (schizophrenia) and  (psychological and mind). Given our research focuses on Bengali literature, we excluded all non-Bengali texts from the data set, which finally contained around 5000 individual posts.

### DSM-5

Reaching an accurate diagnosis is the first step toward appropriately treating any medical condition and mental disorder [24]. The DSM-5 is an authoritative manual that defines and assesses 5 mental disorders. We studied the latest volume of the DSM-5 (2020) to elicit information and improve our understanding of MDD. According to the DSM-5, depression is a condition wherein an individual experiences 5 or more symptoms over the same 2-week period with a pervasive depressed mood or lack of interest and pleasure. The symptoms are a subset of the following:

- Mood swings throughout the majority of the day, practically every day
- Significantly lowered interest in almost all daily activities
- Changes in appetite and notable weight loss or gain
- Diminution in physical movement and a slowing of cognition
- Fatigue or lack of energy nearly every day
- Feelings of worthlessness or excessive and inappropriate guilt on a daily basis
- Reduced ability of concentrating, or indecisiveness, almost every day
- Recurrent thoughts of death, persistent suicidal tendency without a specific plan, or a specific suicide attempt

Furthermore, the DSM-5 underlines the associated features of depression, such as anger, brooding, and compulsive rumination, as well as phobias, excessive concern about physical health,

and complaints of pain—all of which were frequently observed in our data set. The handbook discusses how MDD is correlated with comorbidity and mortality, much of which is attributable to suicide. Suicidal ideation manifests actively in those with depression through words such as “I want to kill myself” or passively through remarks such as “I wish I could simply go to sleep and never wake up” [25]. We discovered identical texts in Bengali in our corpus, “” and “,” as instances of suicidal sentiments.

The latest edition of the DSM-5 added 2 specifiers, the presence of manic symptoms and depression with anxiety distress, to further classify diagnoses. This aided in the precision and concentration of our work.

### Labeling Scheme


Upon analyzing the texts in our data set, we found that the linguistic patterns of users with depression in different stages are consistent with the DSM-5 outlines. We devised a comprehensive labeling technique to categorize the texts into 4 distinct classes based on the duration of suffering, the number of symptoms, the use of absolutist words, suicidal ideation, a mention of manic episodes, and delusional thoughts, among other factors. Given that the research focuses on depression and mental health, we opted to consult with mental health specialists to ensure our approach was sound. For verification, we contacted Ms Tasnuva Huque, who is currently a psychosocial counselor at the Counseling Unit of Brac University, Bangladesh. The labeling technique was revised to strictly adhere to DSM-5 criteria, and finally, Ms Huque authenticated the labeling approach for the data set. Level 4 consists of the most acute and concerning cases on our 4-tiered severity scale, with the weight decreasing for subsequent levels up to level 1, which represents the least problematic instances. In the second stage, we were referred to Ms Syeda Tanzila Huque and Ms Ayesha Seddiqa from the Counseling Unit of Brac University.

The involvement of experts from Brac University ensured sound communication and creation of a labeling guideline. The labels were to be assigned with a number of careful considerations. First, the 4 levels of severity were clearly defined and agreed upon by experts. The remaining part of this section outlines the levels in detail.

Level 4 depression is diagnosed when users' texts contain references to past suicidal attempts or suicidal inclinations and thoughts, self-harm as a result of depression, or diagnosis of
















schizophrenia or borderline personality disorder. Severity level 3 is the broadest category in our data set. It includes texts with references to the need for counseling or medication, postpartum depression or depression during the trimesters, clinical depression associated with psychotic disorders, impaired functioning and phobias (ie, fear of death), lack of appetite, sudden weight loss or gain, delusion, constant mood swings, forgetfulness, breathing difficulties, and other physical health problems. Because this category has a variety of text data and fewer occurrences of each kind, the learning was relatively complex for the models. Severity level 2 consists of written indications of general depression, feelings of hopelessness, loneliness, persistent feelings of instability, and low self-esteem. Lastly, level 1 includes general posts that imply that the users occasionally feel unhappy or that contain mentions of miscellaneous problems that do not pertain to severe depressive symptoms. Some examples of each category are shown in [Table 1](#).






Conflicts during the annotation process arose due to the presence of an array of symptoms belonging to multiple severity levels for a particular item. A user may present hopelessness and elevated levels of frustration along with a statement indicating suicidal ideation. In such cases, the experts resorted to assigning the post the highest level of severity.

Subjective opinion and connotations associated with particular expressions due to their overuse in different contexts presented a problem. Individuals tended to make statements similar to “,” which translates to “What is the point of living?” Such statements are commonplace and are used to express general frustration, hopelessness, and philosophical contemplation. According to annotation guidelines, such statements were to be handled objectively and not taken as casual statements, since they were extracted from mental health platforms and were written by users with depression. Another general rule prohibited the action of inference from user statements and labels could not be assigned based on inferences. When complex differences of opinion occurred, the posts were marked and the differences were resolved with discussion and majority voting. The label generation process involved physical and online sessions in groups of 2 or 3.

The following section elaborates on the experiments conducted on our data set and explores the best-performing architectures. We tested a total of 5 machine learning models and 11 deep learning models.

**Table 1.** Examples in Bengali and their English translations.

Level and examples (Bengali)	Examples (corresponding English translation)
<p><b>Level 4</b></p> <p></p> <p></p> <p></p> <p></p> <p></p>	<p>What is the point of living! It is better to die than live.</p> <p>If suicide was not a sin, then I would commit suicide! Because, there are some problems in the life of some people in this world which cannot be solved without death!</p> <p>I feel like the most depressed person on this planet and I cry a lot. I want to strangle myself sometimes with my own hand and I want to end my life. But I can't. It's very painful. I don't want to live. Please help me out.</p> <p>I've been very depressed for a while and nothing seems to be going well, maybe right now if I committed suicide it would be good.</p> <p>I'm always upset. Sometimes I want to die, but as I said before, there are setbacks.</p>
<p><b>Level 3</b></p> <p></p> <p></p> <p></p> <p></p> <p></p>	<p>I am 6 months pregnant but I can't sleep. I'm suffering from depression.</p> <p>I am suffering from mental depression, what medicine should I take?</p> <p>I have been suffering from depression for a long time. I am also very sick mentally and physically, I feel dizzy all the time, I can't eat anything and I feel nauseous every day.</p> <p>Assalamu Alaikum. I am 10 weeks pregnant. I am very emotional. I cry a lot as a reaction to things that hurt a lot. Whatever the subject may be? Will this be a problem for my baby?</p> <p>I became the mother of the first child through c-section. But I can't be like before. I am suffering from depression. I used to scream and cry. My mom, sister and my husband used to cry outside for me. I just think I'm going to die.</p>
<p><b>Level 2</b></p> <p></p> <p></p> <p></p> <p></p> <p></p>	<p>How does Facebook leave us depressed? I'm very depressed.</p> <p>My hopelessness works all the time and I am in great tension.</p> <p>I suffer from depression most of the time. I have mental problems as a result. What is the remedy?</p> <p>Ways to get rid of frustration and depression. And how to be confident?</p> <p>I feel down all day and lack motivation to get things done. I get very depressed.</p>
<p><b>Level 1</b></p>	

Level and examples (Bengali)	Examples (corresponding English translation)
	I feel like going somewhere far away. Leaving everything for a few days
	I feel depressed almost every day, don't feel good about anything
	I do not feel well at all. I can't focus on anything.
	I feel sad sometimes.
	I don't feel well. I don't like anything.

### Preprocessing

The preprocessing stage included the removal of emojis or emoticons, stopwords, numbers, and other foreign characters, followed by tokenization based on whitespace. A pretrained model, FastText [26], was used for the correction [27] of misspelled words. The length of each post varied from 5 to 300 words. To address the imbalanced nature of the data set, class weights were assigned to give more focus to the minority classes.

### Machine Learning Models

Text feature extractions were performed using the bag-of-words (BoW) and term frequency-inverse document frequency (TF-IDF) methods. BoW is a commonly used, simplifying representation of sentences using word frequency. This model disregards grammar or the order of words, while retaining multiplicity. We used BoW to transform texts into feature vectors. TF-IDF is a statistical model that is often used as a weighting factor for information retrieval, text mining, and user modeling. It evaluates how relevant a word is to a post in a collection of documents. A sparse matrix of representation, based on bigram word counts, of the original posts was obtained using this feature extraction scheme.

We applied a set of SVM and complement NB (a member of the NB family) classifiers to the BoW representations [28] and another set of random forest, logistic regression, and KNN

classifiers to the sparse matrix from the TF-IDF encoding scheme. We incorporated a grid search to find the best set of hyperparameters for each model (ie, C, gamma, and kernel for the SVM classifier; number of estimators for the random forest classifier; and number of neighbors for the KNN).

### Deep Learning Models

Deep learning architectures have achieved groundbreaking results in multiclass classification. We used some standard deep learning architectures and several variations of 1D convolutional layers incorporated with neural network classifiers.

Word embeddings are vector representations of words used as underlying input representations. They generally enhance the performance of sentiment analysis tasks to a great extent. FastText provides word embeddings for 157 languages, including Bengali. Some of the additional benefits of FastText include its extension of the Skip-gram algorithm from Word2vec to create character-level representation of words. In our research, we chose FastText word embedding for the embedding layers. Young et al [29] explored the deep learning trends in text classification. The following section describes how we extracted lower-level sequences from texts and captured long-range dependencies. It also discusses standard deep learning architectures that were used in our experiments.

Figures 1 and 2 show some of the generalized architectures used in our experiments. The experimental design used 1 or a combination of these architectures.

Figure 1. A recurrent neural network (RNN).

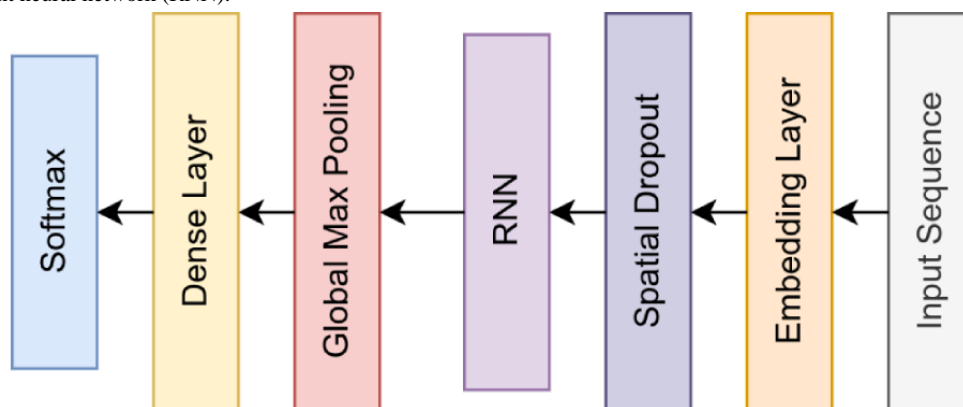
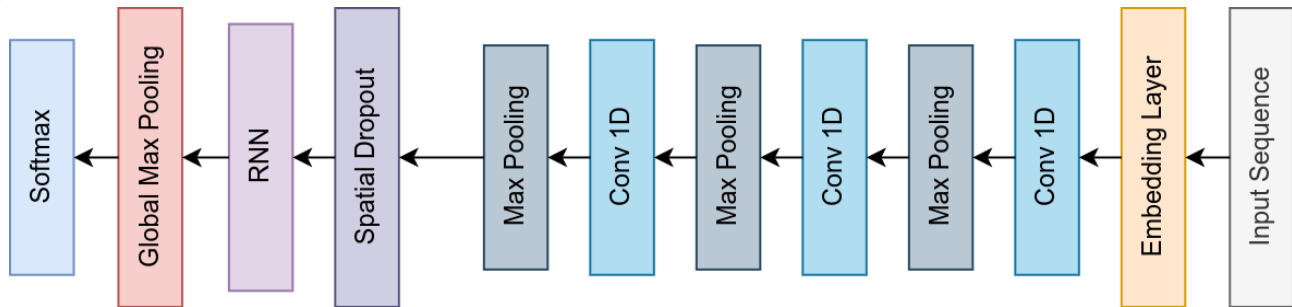


Figure 2. RNN with convolutional blocks. RNN: recurrent neural network.



**BiLSTM**

A BiLSTM connects 2 layers from opposite directions, which enables the architecture to propagate past or future information in both directions. The introduction of LSTM can be traced back to work by Hochreiter and Schmidhuber [30]. The forget, input, and output gates work to capture dependencies and update the contemporary memory cell. The following equations denote the operation of a unidirectional LSTM architecture:

$$\begin{aligned}
 i_t &= \sigma(W_i \times [h_{t-1}, x_t] + b_i) \\
 f_t &= \sigma(W_f \times [h_{t-1}, x_t] + b_f) \\
 q_t &= \tanh(W_q \times [h_{t-1}, x_t] + b_q) \\
 o_t &= \sigma(W_o \times [h_{t-1}, x_t] + b_o) \\
 c_t &= f_t \odot c_{t-1} + i_t \odot q_t \\
 h_t &= o_t \odot \tanh(c_t)
 \end{aligned}$$

The inputs at the current time, the forget gate, and the output gate are represented by  $x_t$ ,  $f_t$ , and  $o_t$ , respectively. The outputs from these gates update the memory cell  $c_t$  and the current hidden state  $h_t$ . The sigmoid function denoted by  $\sigma$  has its domain in the range (0,1). The hyperbolic tangent function has outputs lying between (-1,1). The  $f_t$  function controls how much information is to be retained, and the input gate stores relevant information.

**BiLSTM and Self-Attention**

The Self-Attention layer takes sequences as inputs and outputs aggregate attention scores to find out on which sequences to focus. In our approach, the bidirectional layers were followed by the Self-Attention and GlobalMaxPooling layers. Bahdanau et al [31] proposed state-of-the-art attention architecture that generates context vectors by taking weighted summations of the input vectors and of the hidden cells. Equal emphasis is put on all words of the input sentence, unlike in traditional BiLSTM. In the following equations, the context vector is denoted by  $c_i$  and  $\alpha_{ij}$  refers to the weights that are calculated through backpropagation.  $h_j$  refers to the j-th word in the input sequence.

$$C_i = \sum \alpha_{ij} h_j$$

The weighted sums were calculated for  $t_x$  annotations. Moreover, the weights  $\alpha_{ij}$  were computed using the softmax function. Several researchers have applied attention mechanisms for text classification and reported that the results have exceeded those of simpler architectures [32,33].

**Deep CNN**

CNNs use the concept of sliding a kernel across a tensor to create feature maps. These feature maps capture the important features throughout the text to gain some understanding about the text. The sliding kernel operation on a feature vector over a single channel can be summarized using the following equation:

$$c = f(w^T \times x_{i:h+1} + b)$$

Conneau et al [34] explored deep CNNs and concluded that performance increases with depth. Their architecture comprised convolutional blocks, each convolutional block having 2 convolutional layers along with a batch normalization layer and ReLU nonlinearity. The fully connected layers come after the K-Max Pooling layer. Our architecture comprised 3 1D convolutional layers with 3x3 kernel sizes and 512 dense layer units. The 3 pooling layers used had pool sizes of 3, 5, and 14, respectively. The model was trained with 20 epochs and with a batch size of 32.

**Deep CNN-BiLSTM**

Hassan and Mahmood [35] proposed a convolutional recurrent architecture for sentence classification. They modified the standard CNN-LSTM architecture by excluding the pooling layers. In our experiment, we modified the standard architecture by placing the pooling layers after each of the convolutional layers, as shown in Figure 2. Many authors have compared the CNN-LSTM architecture with stand-alone CNNs, LSTM-CNNs, and other variations [36,37]. In our case, we placed the deep convolutional blocks with the pooling layers. Next, the pooled output, which had a minimal dimension, was passed to an LSTM that learned the ordering of the local features that were extracted.

Experiments were conducted further by modifying the aforementioned architectures with the addition of Self-Attention layers or by changing encoders. The experimental models included a deep CNN-BiGRU, a deep CNN-BiLSTM with Self-Attention, a deep CNN-BiGRU and Self-Attention, and a deep CNN-Self-Attention.

**GRUs**

This architecture excludes the output gate and has fewer parameters. It consists of 2 gates, a reset gate and an update gate. The reset gate deals with the short-term memory of the architecture.

$$\begin{aligned}
 r_t &= \sigma(x_t \times U_r + H_{t-1} \times W_r) \\
 u_t &= \sigma(x_t \times U_u + H_{t-1} \times W_u)
 \end{aligned}$$



The first step of this model involves the computation of candidate hidden states, which is determined by the hidden state of the previous timestamps and multiplied by the reset gate output. The resulting output from the tanh activation function is the candidate hidden state.

$$\hat{H}_t = \tanh(x_t \times U_g + (r_t \cdot H_{t-1}) \times W_g)$$

The extent of information a candidate gate can harbor is determined by the reset gate. The candidate hidden state is then used to calculate the current hidden state. GRUs alone or when used as part of other hybrid architectures have proven to be successful [38,39].

### Pretrained Language Models

The Hugging Face Transformer library offers a variety of pretrained language models [40]. Devlin et al [41] proposed a novel language representation model known as BERT. The model is trained on vast text data to learn bidirectional representations, and the architecture provides room for task-specific fine-tuning. The first part of training involved the implementation of a masked language model. A small proportion of the words were replaced with a fixed token to mask them. The model was trained to predict the masked tokens based on the context. To make the BERT model suitable for classification, a classification token was inserted at the start of the first sentence and a separator token was placed at the end. Additionally, the tokens were assigned a sentence and positional embeddings. A classification layer was placed after the transformer model for emotion detection or sentiment analysis tasks.

### XLM-RoBERTa

This model develops on BERT and uses richer vocabulary for pretraining on multilingual corpora [42,43]. The XLM-RoBERTa architecture used in our experiments comprised a pretrained model trained on an ~3 GB monolingual Bengali corpus [22].

### Evaluations

For our imbalanced classification problem, we used class-weighted evaluation metrics, such as the weighted  $F_1$

score, weighted precision, and recall. For the weighted  $F_1$  score, we adjusted the  $F_1$  scores of each class according to the proportion of samples in that class. The macro- $F_1$  score returned the average  $F_1$  score without considering the number of samples for each class label. Thus, it was insensitive to class imbalance.

$$F_1 \text{ score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

$$\text{Weighted } F_1 \text{ score} = [(N_1 \times \text{Class}_{1\_F1}) + (N_2 \times \text{Class}_{2\_F1}) + (N_3 \times \text{Class}_{3\_F1}) + (N_4 \times \text{Class}_{4\_F1})] / (N_1 + N_2 + N_3 + N_4)$$

## Results

### Traditional Models

We compared the real labels of pure Bengali texts with model predictions and summarized the traditional machine learning models (see Tables 2 and 3). The SVM with a linear kernel achieved the highest generalization ability on TF-IDF vector representations with 78% accuracy. Moreover, this result was marginally better than the SVM model that was trained on the same representations with the radial basis function (rbf) kernel, meaning that the representations are linearly separable to some extent. The following are the values for the hyperparameters that were obtained via a grid search:

BoW:

- kernel: rbf
- C: 55
- gamma: 0.008

TF-IDF:

- kernel: linear
- C: 1

Moreover, the random forest model, which was trained using 25,000 estimators with class weights assigned accordingly to account for the class imbalance, achieved an accuracy of 75%.

**Table 2.** Results with BoW<sup>a</sup> embedding.

Model	Precision	Recall	$F_1$ score	Accuracy
Kernel SVM <sup>b</sup> -rbf <sup>c</sup>	0.73	0.74	0.73	0.74
Kernel SVM-linear	0.71	0.72	0.71	0.72
Complement NB <sup>d</sup>	0.66	0.66	0.66	0.66

<sup>a</sup>BoW: bag-of-words.

<sup>b</sup>SVM: support vector machine.

<sup>c</sup>rbf: radial basis function.

<sup>d</sup>NB: naive Bayes.

**Table 3.** Results with TF-IDF<sup>d</sup> vectorizer.

Model	Precision	Recall	$F_1$ score	Accuracy
Kernel SVM <sup>b</sup> -rbf <sup>c</sup>	0.76	0.77	0.76	0.77
Kernel SVM-linear	0.77	0.78	0.76	0.78
Random forest	0.76	0.75	0.72	0.75
Logistic regression	0.74	0.74	0.74	0.74
KNN <sup>d</sup>	0.70	0.53	0.44	0.53

<sup>a</sup>TF-IDF: term frequency–inverse document frequency.

<sup>b</sup>SVM: support vector machine.

<sup>c</sup>rbf: radial basis function.

<sup>d</sup>KNN: K-nearest neighbor.

### Deep Learning Architectures

We also reported the weighted  $F_1$  scores, macroaverage  $F_1$  scores, and accuracies of our deep learning architectures for pure Bengali texts. Our recurrent models had the following setup: embedding layers, followed by 1D spatial dropout, a stack of recurrent units, and dense layers. The 1D spatial dropout drops the entire feature map; in other words, it drops a feature along with its correlated neighbors by setting its activations to 0. The spatial dropout rate varied from 0.1 to 0.4.

Furthermore, for feature extraction using CNNs, the deep convolutional block was placed after the embedding layer, followed by the 1D spatial dropout. All the models were trained with batch sizes of 32 and 64. The LSTM unit achieved the highest weighted  $F_1$  score and an accuracy of 0.78. The following Table 4 details the results of our final deep learning architectures. The BiGRU reached a weighted  $F_1$  score of 0.81, while the additional layers were able to distinguish the classes

moderately well. The BERT architecture was adjusted with a batch size of 8, a learning rate of  $1 \times 10^{-4}$ , and a fully connected layer consisting of 4096 units with L1 and L2 regularizers set to 0.01. The model was trained for 400 epochs on an NVIDIA RTX 3060 GPU.

Table 5 details the performance of the BiGRU on individual severity levels. It was able to distinguish severity level 4 posts with 81% accuracy. Moreover, severity level 1 and 2 posts could be detected with 86% and 82% accuracy, respectively. It is also important to note that the CNN-based recurrent models achieved higher accuracies in the case of level 4 severity detection. The deep CNN-BiGRU, in particular, achieved 83% accuracy and the deep CNN-BiLSTM yielded 82% accuracy.

The objective of our research was to maximize recall as an indicator since a false-negative case might create a hindrance for a suicidal individual in getting help. The BiGRU achieved a precision of 88%, so it might filter out most of the severe cases if the metric were to be maximized.

**Table 4.** Results of deep learning implementations.

Model	Precision	Recall	Accuracy	$F_1$ score	$F_1$ score (macroaverage)
BiGRU <sup>a</sup>	0.81	0.81	0.81	0.81	0.78
BiLSTM <sup>b</sup> Self-Attention	0.73	0.72	0.72	0.73	0.70
Deep CNN <sup>c</sup> -BiLSTM	0.80	0.77	0.77	0.78	0.76
Deep CNN-BiLSTM Self- Attention	0.77	0.76	0.76	0.76	0.74
BiLSTM	0.77	0.77	0.77	0.77	0.74
BiGRU Self-Attention	0.75	0.74	0.74	0.74	0.73
Deep CNN-BiGRU	0.76	0.76	0.76	0.76	0.74
Deep CNN-BiGRU Self- Attention	0.75	0.73	0.73	0.74	0.73
Deep CNN Self-Attention	0.77	0.77	0.77	0.77	0.75
Monolingual XLM-RoBERTa-BiGRU <sup>d</sup>	0.78	0.78	0.78	0.78	0.75

<sup>a</sup>BiGRU: bidirectional gated recurrent unit.

<sup>b</sup>BiLSTM: bidirectional long short-term memory.

<sup>c</sup>CNN: convolutional neural network.

<sup>d</sup>BERT: bidirectional encoder representations from transformers.

**Table 5.** BiGRU<sup>a</sup> implementation breakdown for each label.

Scale	Precision	Recall	Accuracy	$F_1$ score
Severity level 1	0.85	0.86	0.86	0.86
Severity level 2	0.78	0.82	0.82	0.80
Severity level 3	0.63	0.62	0.62	0.63
Severity level 4	0.88	0.80	0.80	0.84

<sup>a</sup>BiGRU: bidirectional gated recurrent unit.

## Discussion

### Principal Findings

This paper discussed an empirical study that identified the severity of depression using Bengali text-based data. Before categorizing, different cases from classes 1 through 4 were thoroughly studied. The findings suggest that by combining machine learning and deep learning approaches, substantial accuracy may be attained for linguistics data sets on complex psychological tasks, such as the analysis of depression.

### Analysis

In our context, the performance of stand-alone RNNs exceeded expectations due to several reasons. First, composite models, such as BERT, tend to produce average results with small multilabel data sets [44]. Second, the order of words in our data set was significant in concluding the nature of a person's mental state. The stand-alone RNN model captures short-to-medium-range dependencies from input sentences. For example, in the consecutive sentences extracted from the data set "[X]," the anonymous individual writes, "I will die," or "[X]." In the following sentence, they are no longer passive about the issue and state, "[X]," or "I feel like I will probably kill myself someday," indicating a possible suicide attempt in the future. Convolutional and pooling layers tend to disrupt information about the local order of words that must be captured for proper classification. Lastly, in the majority of the cases, our labeling criteria put emphasis on absolutist words, such as "[X]" (mood swing), "[X]" (OCD), "[X]" (hyper), "[X]" (mental issues)' and "[X]" (feel upset all the time) "[X]." The attention model, which assigns attention weights to input representations, requires a substantially larger corpus to accurately calculate which word token is to be assigned a higher weight.

Previous studies performing linguistic analysis on depressive English texts have focused heavily on identification of specific

emotions or mental health issues. In doing so, the authors have performed multilabel classification on a corpus collected from Reddit. Research incorporating Bengali text data is largely limited to classification of depressive and nondepressive texts only. Therefore, the significance of this research lies in demonstrating that deep learning classifiers not only identify specific emotions or conditions but also classify the level of severity. Second, it continues research on Bengali NLP to include classification of hierarchical depressive labels.

The evidence collected from the studies demonstrates that sequential deep learning architectures produce quality results. A proportion of the hybrid architectures suffered from limitations due to unavailability of Bengali language resources. Many state-of-the-art techniques benefit from an abundance of textual data belonging to an array of social topics. However, for categorizing niche social topics in low-resource languages, RNN models generalize better without requiring a large corpus and high computational power.

### Limitations

The lack of pure Bengali texts was a constraint to our work. Bengali-speaking people write texts in Romanized Bengali, which is the representation of Bengali language in English scripts. We were unable to use a fraction of the data initially collected, because some of the texts were code-mixed or written in Romanized Bengali. Moreover, Bengali is a low-resource language and the user posts in our corpus belonged to specialized social topics.

### Conclusion

Future research might focus on designing experiments using Romanized Bengali texts, too. Despite the limitations of the study, our models' overall performance and findings indicate that machine learning and deep learning models are reasonably robust and suitable to identify the severity of mental health conditions.

### Acknowledgments

We are indebted to our supervisor, Dr Md Khalilur Rhaman, for his timely guidance and resources. Furthermore, we would like to express our gratitude toward Ms Tasnuva Huque for her unwavering cooperation. Last but not the least, we appreciate our senior Ajmain Inqiad Alam for his technical guidance.

### Conflicts of Interest

None declared.

## References

1. Torres F. What Is Depression?. URL: <https://www.psychiatry.org/patients-families/depression/what-is-depression> [accessed 2021-12-08]
2. Bachmann S. Epidemiology of suicide and the psychiatric perspective. *Int J Environ Res Public Health* 2018 Jul 06;15(7):1425 [FREE Full text] [doi: [10.3390/ijerph15071425](https://doi.org/10.3390/ijerph15071425)] [Medline: [29986446](https://pubmed.ncbi.nlm.nih.gov/29986446/)]
3. World Health Organization. Depression. URL: <https://www.who.int/news-room/fact-sheets/detail/depression> [accessed 2021-12-08]
4. Barney LJ, Griffiths KM, Jorm AF, Christensen H. Stigma about depression and its impact on help-seeking intentions. *Aust N Z J Psychiatry* 2006;40(1):51-54 [FREE Full text] [doi: [10.1080/j.1440-1612006.01741.x](https://doi.org/10.1080/j.1440-1612006.01741.x)]
5. Losada DE, Crestani F, Parapar J. eRisk 2020: self-harm and depression challenges. In: *Advances in Information Retrieval. ECIR 2020. Lecture Notes in Computer Science, Vol 12036*. Cham: Springer; 2020.
6. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960 Feb 01;23(1):56-62 [FREE Full text] [doi: [10.1136/jnnp.23.1.56](https://doi.org/10.1136/jnnp.23.1.56)] [Medline: [14399272](https://pubmed.ncbi.nlm.nih.gov/14399272/)]
7. Ellendorff T, Foster S, Rinaldi F. The PsyMine Corpus - a corpus annotated with psychiatric disorders and their etiological factors. 2016 Presented at: Tenth International Conference on Language Resources and Evaluation (LREC'16); May 23-28, 2016; Portorož, Slovenia.
8. Alonso SG, de la Torre-Díez I, Hamrioui S, López-Coronado M, Barreno DC, Nozaleda LM, et al. Data mining algorithms and techniques in mental health: a systematic review. *J Med Syst* 2018 Jul 21;42(9):161. [doi: [10.1007/s10916-018-1018-2](https://doi.org/10.1007/s10916-018-1018-2)] [Medline: [30030644](https://pubmed.ncbi.nlm.nih.gov/30030644/)]
9. Yates A, Cohan A, Goharian N. Depression and self-harm risk assessment in online forums. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics; 2017:2968-2978.
10. MacAvaney S, Desmet B, Cohan A, Soldaini L, Yates A, Zirikly A, et al. RSDD-Time: temporal annotation of self-reported mental health diagnoses. arXiv. Preprint posted online June 20, 2018 [FREE Full text] [doi: [10.48550/arXiv.1806.07916](https://doi.org/10.48550/arXiv.1806.07916)]
11. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng J* 2014 Dec;5(4):1093-1113. [doi: [10.1016/j.asej.2014.04.011](https://doi.org/10.1016/j.asej.2014.04.011)]
12. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJP, Dobson RJB, et al. Characterisation of mental health conditions in social media using informed deep learning. *Sci Rep* 2017 Mar 22;7:45141 [FREE Full text] [doi: [10.1038/srep45141](https://doi.org/10.1038/srep45141)] [Medline: [28327593](https://pubmed.ncbi.nlm.nih.gov/28327593/)]
13. Mageed M, Ungar L. EmoNet: fine-grained emotion detection with gated recurrent neural networks. 2017 Presented at: 55th Annual Meeting of the Association for Computational Linguistics; July 2017; Vancouver, Canada p. 718-728. [doi: [10.18653/v1/p17-1067](https://doi.org/10.18653/v1/p17-1067)]
14. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. 2016 Presented at: 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2016; San Diego, CA p. 1480-1489. [doi: [10.18653/v1/n16-1174](https://doi.org/10.18653/v1/n16-1174)]
15. Ive J, Gkotsis G, Dutta R, Stewart R, Velupillai S. Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health. 2018 Presented at: Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic; June 2018; New Orleans, LA p. 69-77. [doi: [10.18653/v1/w18-0607](https://doi.org/10.18653/v1/w18-0607)]
16. Cohan A, Desmet B, Yates A, Soldaini L, MacAvaney S, Goharian N. SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions. arXiv. Preprint posted online June 13, 2018 [FREE Full text]
17. Mustafa R, Ashraf N, Ahmed FS, Ferzund J, Shahzad B, Gelbukh A. A multiclass depression detection in social media based on sentiment analysis. In: Latifi S, editor. *17th International Conference on Information Technology–New Generations (ITNG 2020)*. *Advances in Intelligent Systems and Computing*, Vol 1134. Cham: Springer; 2020.
18. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. 1992 Presented at: COLT92: Fifth Annual Workshop on Computational Learning Theory; July 27-29, 1992; Pittsburgh, PA. [doi: [10.1145/130385.130401](https://doi.org/10.1145/130385.130401)]
19. Al-Garadi M, Yang YC, Cai H, Ruan Y, O'Connor K, Graciela GH, et al. Text classification models for the automatic detection of nonmedical prescription medication use from social media. *BMC Med Inform Decis Mak* 2021 Jan 26;21(1):27 [FREE Full text] [doi: [10.1186/s12911-021-01394-0](https://doi.org/10.1186/s12911-021-01394-0)] [Medline: [33499852](https://pubmed.ncbi.nlm.nih.gov/33499852/)]
20. Uddin AH, Bapery D, Arif ASM. Depression analysis from social media data in bangla language using long short term memory (LSTM) recurrent neural network technique. 2019 Presented at: 5th International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2); July 11-12, 2019; Rajshahi, Bangladesh p. 1-4. [doi: [10.1109/ic4me247184.2019.9036528](https://doi.org/10.1109/ic4me247184.2019.9036528)]
21. Khan MRH, Afroz US, Masum AKM, Abujar S, Hossain SA. A deep learning approach to detect depression from Bengali text. In: Hassanien AE, Bhattacharyya S, Chakrabati S, Bhattacharya A, Dutta S, editors. *Emerging Technologies in Data Mining and Information Security*. *Advances in Intelligent Systems and Computing*, Vol 1300. Singapore: Springer; 2021.
22. Hugging Face. Indic-Transformers Bengali XLMRoBERTa. URL: <https://huggingface.co/neuralspace-reverie/indic-transformers-bn-xlmroberta> [accessed 2022-09-17]
23. omanwhatiscomputer / depression-severity. URL: <https://github.com/omanwhatiscomputer/depression-severity> [accessed 2022-09-17]

24. Rachel AD. DSM-5 handbook of differential diagnosis. *Am J Psychiatry* 2014;171(5):589-590.
25. Truschel J. Depression Definition and DSM-5 Diagnostic Criteria. URL: <https://www.psycom.net/depression-definition-dsm-5-diagnostic-criteria/> [accessed 2021-12-08]
26. Grave E, Bojanowski P, Gupta P, Joulin A, Mikolov T. Learning word vectors for 157 languages. 2018 Presented at: 11th International Conference on Language Resources and Evaluation (LREC 2018); May 7-12, 2018; Miyazaki, Japan.
27. Kaiser A. Bengali Word Spelling Correction Using Pre-trained Word2Vec. URL: <https://medium.com/analytics-vidhya/bengali-word-spelling-correction-using-pre-trained-word2vec-1f9938f48b09> [accessed 2021-06-08]
28. Zhang Y, Jin R, Zhou Z. Understanding bag-of-words model: a statistical framework. *Int J Mach Learn Cyber* 2010 Aug 28;1(1-4):43-52 [FREE Full text] [doi: [10.1007/s13042-010-0001-0](https://doi.org/10.1007/s13042-010-0001-0)]
29. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. arXiv. Preprint posted online August 9, 2017 [FREE Full text]
30. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997 Nov 15;9(8):1735-1780 [FREE Full text] [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)] [Medline: [9377276](https://pubmed.ncbi.nlm.nih.gov/9377276/)]
31. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. 2015 Presented at: 3rd International Conference on Learning Representations (ICLR 2015); May 7-9, 2015; San Diego, CA.
32. Jing R. A self-attention based LSTM network for text classification. *J Phys: Conf Ser* 2019 Apr 26;1207:012008. [doi: [10.1088/1742-6596/1207/1/012008](https://doi.org/10.1088/1742-6596/1207/1/012008)]
33. Li W, Qi F, Tang M, Yu Z. Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing* 2020 Apr;387:63-77. [doi: [10.1016/j.neucom.2020.01.006](https://doi.org/10.1016/j.neucom.2020.01.006)]
34. Conneau A, Schwenk H, Barrault L, Lecun Y. Very deep convolutional networks for text classification. 2017 Presented at: 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers; April 2017; Valencia, Spain p. 1107-1116. [doi: [10.18653/v1/e17-1104](https://doi.org/10.18653/v1/e17-1104)]
35. Hassan A, Mahmood A. Convolutional recurrent deep learning model for sentence classification. *IEEE Access* 2018;6:13949-13957. [doi: [10.1109/ACCESS.2018.2814818](https://doi.org/10.1109/ACCESS.2018.2814818)]
36. Sosa PM. Twitter Sentiment Analysis using combined LSTM-CNN Models. 2017. URL: <https://konukoi.com/blog/2018/02/19/twitter-sentiment-analysis-using-combined-lstm-cnn-models/> [accessed 2022-09-17]
37. Zhou C, Sun C, Liu Z, Lau F. A C-LSTM neural network for text classification. arXiv. Preprint posted online November 27, 2015 [FREE Full text]
38. Yan W, Zhou L, Qian Z, Xiao L, Zhu H. Sentiment analysis of student texts using the CNN-BiGRU-AT model. *Sci Prog* 2021 Oct 22;2021:1-9 [FREE Full text] [doi: [10.1155/2021/8405623](https://doi.org/10.1155/2021/8405623)]
39. Zhou L, Bian X. Improved text sentiment classification method based on BiGRU-Attention. *J Phys: Conf Ser* 2019 Nov 01;1345(3):032097. [doi: [10.1088/1742-6596/1345/3/032097](https://doi.org/10.1088/1742-6596/1345/3/032097)]
40. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: state-of-the-art natural language processing. 2020 Presented at: 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP 2020); November 8-12, 2020; Online. [doi: [10.18653/v1/2020.emnlp-demos.6](https://doi.org/10.18653/v1/2020.emnlp-demos.6)]
41. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online October 11, 2018 [FREE Full text]
42. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen S, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv. Preprint posted online July 26, 2019 [FREE Full text] [doi: [10.48550/arXiv.1907.11692](https://doi.org/10.48550/arXiv.1907.11692)]
43. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzm F, et al. Unsupervised cross-lingual representation learning at scale. arXiv. Preprint posted online November 5, 2019 [FREE Full text]
44. Ezen-Can A. A comparison of LSTM and BERT for small corpus. arXiv. Preprint posted online September 11, 2020 [FREE Full text]

## Abbreviations

- API:** Application Programming Interface
- BERT:** bidirectional encoder representations from transformers
- BiGRU:** bidirectional gated recurrent unit
- BiLSTM:** bidirectional long short-term memory
- BoW:** bag-of-words
- CNN:** convolutional neural network
- KNN:** K-nearest neighbor
- LIWC:** Linguistic Inquiry and Word Count
- LSTM:** long short-term memory
- MDD:** major depressive disorder
- NB:** naive Bayes
- NLP:** natural language processing
- OCD:** obsessive-compulsive disorder

**rbf:** radial basis function

**RNN:** recurrent neural network

**SVM:** support vector machine

**TF-IDF:** term frequency–inverse document frequency

*Edited by A Mavragani; submitted 02.01.22; peer-reviewed by T Basu, M Rodrigues; comments to author 22.03.22; revised version received 16.05.22; accepted 20.07.22; published 28.09.22.*

*Please cite as:*

*Kabir MK, Islam M, Kabir ANB, Haque A, Rhaman MK*

*Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques*

*JMIR Form Res 2022;6(9):e36118*

*URL: <https://formative.jmir.org/2022/9/e36118>*

*doi: [10.2196/36118](https://doi.org/10.2196/36118)*

*PMID: [36169989](https://pubmed.ncbi.nlm.nih.gov/36169989/)*

©Muhammad Khubayeeb Kabir, Maisha Islam, Anika Nahian Binte Kabir, Adiba Haque, Md Khalilur Rhaman. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 28.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Comparing Transactional eHealth Literacy of Individuals With Cancer and Surrogate Information Seekers: Mixed Methods Study

Taylor S Vasquez<sup>1\*</sup>, BSc, MA; Carma L Bylund<sup>1,2\*</sup>, PhD; Jordan Alpert<sup>1\*</sup>, PhD; Julia Close<sup>3\*</sup>, MD; Tien Le<sup>1\*</sup>, BSc; Merry Jennifer Markham<sup>3\*</sup>, MD; Greenberry B Taylor<sup>4\*</sup>, PhD; Samantha R Paige<sup>2,5\*</sup>, PhD

<sup>1</sup>College of Journalism and Communications, University of Florida, Gainesville, FL, United States

<sup>2</sup>College of Medicine, University of Florida, Gainesville, FL, United States

<sup>3</sup>Division of Hematology & Oncology, College of Medicine, University of Florida, Gainesville, FL, United States

<sup>4</sup>Department of Communication, Flagler College, St. Augustine, FL, United States

<sup>5</sup>Johnson & Johnson, Inc, Health & Wellness Solutions, New Brunswick, NJ, United States

\* all authors contributed equally

**Corresponding Author:**

Taylor S Vasquez, BSc, MA

College of Journalism and Communications

University of Florida

1885 Stadium Road

Gainesville, FL, 32608

United States

Phone: 1 3215370306

Email: [tsthelander@ufl.edu](mailto:tsthelander@ufl.edu)

## Abstract

**Background:** The number of adults entering higher-risk age groups for receiving a cancer diagnosis is rising, with predicted numbers of cancer cases expected to increase by nearly 50% by 2050. Living with cancer puts exceptional burdens on individuals and families during treatment and survivorship, including how they navigate their relationships with one another. One role that a member of a support network may enact is that of a surrogate seeker, who seeks information in an informal capacity on behalf of others. Individuals with cancer and surrogate seekers often use the internet to learn about cancer, but differences in their skills and strategies have received little empirical attention.

**Objective:** This study aimed to examine the eHealth literacy of individuals with cancer and surrogate information seekers, including an investigation of how each group evaluates the credibility of web-based cancer information. As a secondary aim, we sought to explore the differences that exist between individuals with cancer and surrogate seekers pertaining to eHealth literacies and sociodemographic contexts.

**Methods:** Between October 2019 and January 2020, we conducted a web-based survey of 282 individuals with cancer (n=185) and surrogate seekers (n=97). We used hierarchical linear regression analyses to explore differences in functional, communicative, critical, and translational eHealth literacy between individuals with cancer and surrogate seekers using the Transactional eHealth Literacy Instrument. Using a convergent, parallel mixed methods design, we also conducted a thematic content analysis of an open-ended survey response to qualitatively examine how each group evaluates web-based cancer information.

**Results:** eHealth literacy scores did not differ between individuals with cancer and surrogate seekers, even after adjusting for sociodemographic variables. Individuals with cancer and surrogate seekers consider the credibility of web-based cancer information based on its channel (eg, National Institutes of Health). However, in evaluating web-based information, surrogate seekers were more likely than individuals with cancer to consider the presence and quality of scientific references supporting the information. Individuals with cancer were more likely than surrogate seekers to cross-reference other websites and web-based sources to establish consensus.

**Conclusions:** Web-based cancer information accessibility and evaluation procedures differ among individuals with cancer and surrogate seekers and should be considered in future efforts to design web-based cancer education interventions. Future studies may also benefit from more stratified recruitment approaches and account for additional contextual factors to better understand the unique circumstances experienced within this population.

**KEYWORDS**

eHealth literacy; cancer communication; individuals with cancer; surrogate seekers; web-based information credibility appraisal

## Introduction

### Overview

The number of adults entering higher-risk age groups for receiving a cancer diagnosis is rising, with predicted numbers of cancer cases expected to increase by nearly 50% by 2050 [1]. Living with cancer puts exceptional burden on individuals and families during treatment and survivorship, including how they navigate their relationships with one another [2]. To best support individuals with cancer and members of their support networks with evidence-based programs, a better understanding of the unique needs and roles of each group is warranted [3].

Active participation in health care decision-making leads to better outcomes and increased quality of life and helps patients receive appropriate and cost-effective treatments [4]. Unfortunately, support networks tend to underassess their quality of life, bringing into question the understanding this population has of the perspective of an individual with cancer [5]. Informed support networks are better able to provide productive support, assist in treatment compliance, and help ensure the continuity of care for the patient [6]. Similarly, well-informed individuals with cancer who participate in shared decision-making with their clinicians have more positive cognitive outcomes related to their care [7,8].

One way in which individuals with cancer and support networks serve active roles in health care experiences is web-based health information seeking [9]. Blogs and social media have evolved past 1-way information websites into expansive, 2-way communicative resources that help individuals with cancer and caregivers connect and learn from other people's experiences and expertise [10]. The availability of health and medical information through web-based forums and sources has increased dramatically over the last several decades and has created an abundance of opportunities for patients and support networks to engage in seeking health information [11,12]. Members of a support network who search for information pertaining to their family member or friend's health and diagnosis are called *surrogate seekers* [11]. The act of surrogate seeking is defined as seeking "information in a nonprofessional or informal capacity on behalf (or because) of others without necessarily being asked to do so" [13]. Although these groups frequently use the internet to search for health information, the general population unfortunately has a challenging time evaluating the quality and veracity of web-based health information [14], and research has demonstrated that this challenge is evident among patients living with chronic ailments such as cancer [10]. Web-based cancer misinformation is prevalent and has the potential to cause harm to its consumers [15]. Investigating how individuals with cancer and surrogate seekers navigate web-based cancer information is imperative to begin developing interventions that support each group in health decision-making.

### Transactional eHealth Literacy

eHealth literacy is a dynamic, intrapersonal skill set that is shaped by the experiences, technologies, and opportunities available to an individual at a given time [16]. Paige et al [17] defined transactional eHealth literacy as, "the ability to locate, understand, exchange, and evaluate health information from the internet in the presence of dynamic contextual factors, and to apply the knowledge gained for the purposes of maintaining or improving health." eHealth literacy is an expanding field of research, as a large portion of health-related messages and information is circulated and accessed through web-based, media, and digital sources [18]. eHealth literacy requires combining both knowledge and skills from a diverse set of domains and is inherently and increasingly relevant to scholars, patients, and other individuals [18]. The Transactional Model of eHealth Literacy (TMeHL) centralizes the concept of communication within the context of how web-based health information is accessed, evaluated, and applied to inform health decisions. The TMeHL functions under a broad assumption that the ability of an individual to counteract challenges during the web-based experience is a continuous process, and the process is constantly modified according to diverse eHealth contextual factors and prior eHealth experiences. Transactional eHealth literacy includes the following four competencies: (1) functional (ie, the ability to locate and understand web-based information); (2) communicative (ie, the ability to exchange information among individuals within web-based contexts); (3) critical (ie, the ability to appraise and evaluate the source and content of information found on the web); and (4) translational (ie, the ability to use information learned from the internet to inform health care decisions) skills [19]. eHealth literacy is associated with sociodemographic and psychosocial variables. For example, people who are younger, report higher education, and use electronic devices more frequently have higher eHealth literacy than their counterparts [17,20]. A systematic review found that caregivers generally have higher health literacy levels than patients [21], but the eHealth literacy skills of patients with cancer and caregivers have received less attention. Caregivers are more likely to be female and younger, so differences are assumed [22], but confirmatory research is needed.

Sillence et al [23] reported that patients and caregivers have different web-based information needs and uniquely engage with website content to determine its quality and relevance. A report by the Pew Research Center [24] found that caregivers are more likely to consult web-based rankings of clinicians and medical facilities and use web-based reviews of drugs and medical treatments, whereas patients seek information about their diagnosis, the causes and spread of cancer, treatment options, and the side effects of treatment. There remains a lack of evidence regarding the specific cues and processes that patients and surrogate seekers use to evaluate the credibility, or the trust and quality, of web-based health information.



## Critical eHealth Literacy

Critical eHealth literacy is defined as the knowledge and ability of a person to evaluate the credibility, relevance, and risk of exchanging web-based health information [17]. The perceived credibility of health information is associated with the recipient's satisfaction with the information found and is ultimately linked to proactive behavior change [23,25]. There is much debate about how to conceptualize credibility [26], but it is generally dependent on whether a consumer considers a piece of information to be believable or guided by trustworthiness and expertise [27]. When evaluating web-based health information, the public is encouraged to consider the source of the information and rely on educational or government agencies to cross-reference the content with other reliable sources and to consider the date on which the content was published, among other indicators [28]. Examining how patients and surrogate seekers assess the credibility of web-based cancer information could provide valuable insights into how lay audiences operationalize the concept of credibility. Such evidence will be valuable for the design and consistent evaluation of patient education resources that are developed to be perceived as credible by their recipients. Furthermore, this evidence will inform how to best deliver education from diverse sources within web-based contexts and offered by offline contacts (eg, clinicians and support networks).

## Purpose

This study aimed to evaluate the eHealth literacy of individuals with cancer and surrogate seekers. Given that perceived skills to evaluate web-based health information do not always translate into proficient performance behaviors [29], we also asked individuals with cancer and surrogate seekers to reflect on the processes they use to evaluate web-based cancer information.

- Hypothesis 1: compared with individuals with cancer, surrogate seekers will have a higher self-reported ability related to functional, communicative, critical, and translational eHealth literacy.
- Research question 1: What differences exist between individuals with cancer and surrogate seekers pertaining to eHealth literacy and sociodemographic contexts?
- Research question 2: What processes do individuals with cancer and surrogate seekers use to evaluate the credibility of web-based cancer information?

## Methods

### Recruitment

Between October 2019 and January 2020, we conducted a 20-minute web-based survey with adults registered with the broad consent research registry of a large southeastern medical university. The broad consent research registry is a database of individuals with cancer who volunteered to be contacted about research opportunities. We were provided with contact information for individuals with cancer who had consented to be contacted for research and who also had an International Classification of Diseases-10 code identifying cancer. Identified individuals with cancer (N=6847) were sent an email invitation, receiving up to 1 reminder. All identified individuals with cancer

received a follow-up email because of our not distributing individualized survey links. This allowed potential participants to forward the email to other eligible individuals. The eligibility criteria included being an English-speaking adult (aged  $\geq 18$  years) and having used the internet to look for advice or information about (1) their own cancer or (2) a family member or friend's cancer in the past 6 months. Surrogate seekers were identified through a snowball sampling technique. We sent an email invitation to the identified individuals with cancer:

*If you have not searched for cancer information in the past six months but have a family member or friend that searches online cancer information for you, we would like to hear from them. Please consider forwarding them this email.*

We did not identify dyads of patients and their support networks to examine similarities, differences, or trends. This study recruited surrogate seeker participants through referrals from individuals with cancer who completed the survey, but we did not establish any dyadic connections between patients and the surrogate seekers they recommended to complete the survey.

The participants who completed the survey were remunerated with a US \$25 e-gift card for their time. The study data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) tools hosted at the University of Florida [30]. REDCap is a secure, web-based application designed to support data capture for research studies. This manuscript reports a secondary analysis of items from the larger survey.

### Measures

The sociodemographic characteristics for this study, including age, sex, race, and socioeconomic status, were measured using items adopted from the Health Information National Trends Survey and the US Census Bureau (Multimedia Appendix 1). Participants indicated whether they were a person living with their own cancer diagnosis (ie, a patient). If not affirmed, participants were asked if they had used the internet to look for advice or information about someone else's cancer. If they said yes, they were included in the study as a surrogate information seeker.

We measured eHealth literacy using the Transactional eHealth Literacy Instrument (TeHLI) of Paige et al [17]. The TeHLI previously underwent a rigorous instrument development and testing procedure, yielding 4 dimensions consistent with the TMeHL [19]. The TeHLI was selected for this study because of its focus on transactional exchanges within the eHealth domain and because it extends beyond other previously established eHealth measures such as the eHealth Literacy Scale (eHEALS) [17]. Although eHEALS is popular and psychometrically sound [19,31], scholars have noted its limitations in content validity informed by the evolving social possibilities created by eHealth technology, in addition to the lack of correlation between eHEALS scores and enacted task performance pertaining to web-based health information seeking [17]. It also does not sufficiently address the critical and communicative skills included in accurately assessing eHealth literacy [32]. Rather than acting as a static competency, eHealth

is fluid and dynamic in nature, which allows individuals opportunities to manage communicative interactions in numerous contexts from multiple sources [33]. This ability to appraise, evaluate, and exchange information between web-based sources is inherently transactional, and this continuous process influences individuals' relational and cultural contexts and acts as an arbiter of social exchange [17,34]. Norman [32] recommended that eHealth literacy instruments should expand beyond the eHEALS to measure the transactional features afforded by eHealth, and the TeHLI accomplishes this.

The multidimensional TeHLI includes 18 items anchored on a 5-point Likert-type scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The instrument measures four competencies: (1) functional (eg, "I can summarize basic health information from the internet in my own words"); (2) communicative (eg, "I have the skills I need to talk about health topics on the internet with multiple users at the same time"); (3) critical (eg, I can tell when health information on the internet is fake"); and (4) translational (eg, "I can use the internet as a tool to improve my health"). The internal consistency of data from each dimension was sufficient for patients (Cronbach  $\alpha$ =.85-.87) and surrogate seekers (Cronbach  $\alpha$ =.83-.91). We also sought to understand how patients and surrogate seekers appraise web-based health information. A single, open-ended item was included in the survey asking, "When you found information online, how did you decide if the information was credible?"

### Statistical Analysis

This mixed methods study used a convergent parallel design, in which the quantitative and qualitative data collection occurred concurrently [35]. Per the convergent parallel design, we used the quantitative and qualitative results to compare findings to yield a more holistic understanding of the data [35]. Convergent designs aim to obtain different yet complementary data on the same topic, assist in producing rigorous scholarship through the independent collection of data, and enhance the subsequent comparison and integration of the results of each method. We generated a series of descriptive statistics to summarize the sociodemographic characteristics of the individuals with cancer and caregivers who participated in the survey. We also conducted a 2-tailed independent samples *t* test to examine if the age differed by group, and we conducted chi-square analyses to detect group differences in sex, race, education, and marital

status. To test hypothesis 1, we conducted 4 hierarchical multiple linear regression analyses. In step 1, we entered sociodemographics that were empirically shown to correlate with each eHealth literacy competency. In step 2, we estimated whether a difference in each eHealth literacy score existed between individuals with cancer and surrogate seekers while adjusting for sociodemographics in step 1. Variance inflation factors and collinearity were also monitored. Statistical significance was detected at  $P$ =.05 or lower, and we reported 95% CIs.

As part of the survey, an open-ended question (When you found information online, how did you decide if the information was credible?) was included to investigate how participants determine whether the health information found on the web was credible. Using a content analysis method, we first completed an inductive open-coding process to determine common themes among both respondent populations [36]. Through this procedure, the identification of keywords and commonly discussed content assisted in the development of 8 codes. To understand which action words (or operational behaviors and skills) were used by participants to evaluate web-based cancer information, we extracted the verbs used in their self-reflective procedure. We compared these action words to those reported in a concept analysis of eHealth literacy definitions, models, and measures to inform the most recent definition of eHealth literacy [17]. Table 1 includes the codebook, which went through several training iterations with 2 researchers who coded independently. An acceptable level of intercoder reliability was established (Cohen  $\kappa$ =0.73) after 2 coders independently coded 20% of the data set. A single coder evaluated the remaining data set and determined coder consensus when necessary. A series of chi-square analyses was conducted to examine differences in credibility appraisal (ie, determined source credibility and determined channel credibility) of web-based health information according to patient and surrogate seeker status.

As a final step, integration of the data occurred by merging the quantitative results with the qualitative results [35]. Through triangulation, we compared the results of each data set, quantitative and qualitative, to draw the conclusions set forth in the results presented [37].

**Table 1.** Codes, operational definitions, and examples to guide open-ended responses.

Code	Definition	Respondent quotes
Determined channel credibility	This code should be used when the respondent answers the question by saying they looked for credible websites and paid attention to where the information was coming from such as Mayo Clinic or a .gov or .edu website. If the respondent double-checked information, read reviews, or trusted the website, these answers could fall under this category as well.	<ul style="list-style-type: none"> <li>Published in a reputable peer-reviewed journal</li> <li>It came from a credible source like a journal or medical center</li> <li>Study was done by a credible medical institution</li> <li>Research or credible sample size or investigators</li> <li>By whom it was provided</li> <li>I made sure the website is accredited</li> </ul>
Determined source credibility	This code should be used when the respondent answers the question by saying they looked up to see who the author was and if they were credible.	<ul style="list-style-type: none"> <li>I noted the author and the credibility of the institution which it represented</li> <li>I had to x-ref dates and study authors to see what was most current, who was still strong in the field, etc.</li> <li>I looked at the credentials of the author</li> </ul>
Checked citations for scientific support	This code should be used when the respondent answers the question by saying they checked the resources of the information to see if it was coming from a credible source. An example of this could be if the respondent looks at the resource the information was taken from, checked the sources at the bottom of the website, age of the article, etc.	<ul style="list-style-type: none"> <li>I looked at the resources the information came from</li> <li>I checked the sources at the bottom of the website</li> <li>Age of article references</li> <li>Researched the sources listed</li> <li>Looked for citations from doctors at the bottom</li> <li>Source references</li> </ul>
Cross-referenced content with other web-based sources	This code should be used when the respondent answers the question by saying they looked at or researched several different sources or web pages to determine if the web-based health information was consistent with each other.	<ul style="list-style-type: none"> <li>How frequently it was repeated in both articles and university-based publications</li> <li>Cross-referencing multiple sites</li> <li>I checked additional websites to compare information as accurate</li> </ul>
Cross-referenced content with recommendations from health care	This code should be used when the respondent answers the question by saying they researched health information on the web to confirm that it aligned with what their clinician recommended. This code can also be used when a patient or caregiver confirms information found on the web between 2 websites.	<ul style="list-style-type: none"> <li>Compared information with what I had gotten from my medical source</li> <li>Seemed like it was in line with what my health care clinician told me</li> <li>Compared with information I received from my physician and medical team</li> </ul>
Discussed content with a health care clinician	This code should be used when the respondent answers the question by saying they discussed the information they found on the web with their health care clinician to check its credibility.	<ul style="list-style-type: none"> <li>I reviewed my symptoms with the information provided, then discussed the symptoms and information with my physician</li> <li>I asked my doctor about it</li> </ul>
Miscellaneous	This code should be used when the respondent answers the question with a response that does not fit into any other predefined reason or cannot be explicitly categorized.	<ul style="list-style-type: none"> <li>Yes it seemed helpful and made me make a doctor appointment</li> <li>Asked a family member</li> <li>It sounded reasonable with what I knew already</li> <li>I trust hospital information</li> </ul>
Uncodeable	This code should be used when the respondent answers the question with a response that does not pertain to the information asked. An example of this could be they did not answer the question correctly or provided information that is not relevant to this data.	<ul style="list-style-type: none"> <li>Not sure</li> <li>Yes I did</li> <li>Yes</li> <li>Had way to know just had to trust reputation</li> <li>Looks real if all of them about what I think</li> </ul>

## Ethical Considerations

This study received institutional review board approval from the University of Florida (IRB#201802322). Each participant completed a waiver of informed consent providing them with clear expectations of what the study entailed and provided their consent before participating. Any identifiable respondent information was anonymized according to ethical privacy standards.

## Results

### Sample Characteristics

A total of 303 participants responded to the survey. A small proportion ( $n=21$ , 6.9%) of the participants reported not searching for web-based cancer information in the past 6 months and were excluded from the analyses. The final sample consisted of 282 participants, which included individuals with cancer ( $n=185$ , 65.6%) and surrogate seekers ( $n=97$ , 34.4%).

Tables 2 and 3 show the sociodemographics of the sample, separated into individuals with cancer and surrogate seeker

groups. Individuals with cancer and surrogate seekers were predominantly White (individuals with cancer: 142/185, 76.7% and surrogate seekers: 69/97 71%), female (individuals with cancer: 105/185, 56.8% and surrogate seekers: 65/97, 67%), college educated (individuals with cancer: 100/185, 54.1% and surrogate seekers: 58/97, 59%), married (individuals with cancer: 105/185, 56.8% and surrogate seekers: 47/97, 48%), and from the southeastern region of the nation (individuals with cancer: 143/185, 77.3% and surrogate seekers: 77/97, 79%). The sample was also predominantly non-Hispanic, with only 7 (2.5%) participants identifying as Hispanic. The most common individual with cancer surrogate seekers identified themselves as seeking information for was either a friend (20/97, 21%) or a spouse (13/97, 13%). Multimedia Appendix 2 shows the surrogate seeker types in detail. Age (mean 55.47, SD 15.10; range 20-88 years) was the only sociodemographic variable that varied according to patient and surrogate seeker status. An independent sample 2-tailed  $t$  test showed that individuals with cancer reported an older age (mean 57.44, SD 14.29 years) than surrogate seekers in this sample (mean 51.72, SD 16.0;  $t_{234}=2.80$ ;  $P=.01$ , 95% CI 1.70-9.75).

**Table 2.** Sociodemographic characteristics (patient sample, N=185).

Variables	Values
Age, mean (SD)	57.44 (14.29)
<b>Sex, n (%)</b>	
Male	50 (27)
Female	105 (56.8)
Intersex	1 (0.5)
Missing	29 (15.7)
<b>Race, n (%)</b>	
White	142 (76.8)
Hispanic or Latino	4 (2.2)
Black or African American	8 (4.3)
Other	2 (1.1)
Missing	29 (15.7)
<b>Education, n (%)</b>	
Less than high school	2 (1.1)
High school or General Education Development	15 (8.1)
Some college	32 (17.3)
Completed college	45 (24.3)
Completed some postgraduate	10 (5.4)
Master's degree	31 (16.8)
Other advanced degree beyond master's	14 (7.6)
Missing	36 (19.5)
<b>Marital status, n (%)</b>	
Single	43 (23.2)
Partnered	9 (4.9)
Married	105 (56.8)
Missing	67 (36.2)
<b>Geographic region, n (%)</b>	
Midwest	0 (0)
Northeast	7 (3.8)
Southeast	143 (77.3)
Southwest	6 (3.2)
West	0 (0)
Missing	29 (15.7)
<b>Type of cancer, n (%)<sup>a</sup></b>	
Breast	39 (26.4)
Skin (squamous cell carcinoma, basal cell carcinoma, melanoma, and Merkel cell)	26 (17.6)
Blood (leukemia, lymphoma, and myeloma)	12 (6.6)
Lung	10 (6.8)
Thyroid	8 (5.4)
Prostate	6 (4.1)
Colon and rectal	5 (3.4)
Non-Hodgkin lymphoma	4 (2.7)

Variables	Values
Ovarian	4 (2.7)
Pancreatic	4 (2.7)
Endometrial	2 (1.4)
Esophageal	2 (1.3)
Head and neck	2 (1.4)
Kidney	2 (1.3)
Mesothelioma	2 (1.4)
Parotid gland	2 (1.3)
Throat	2 (1.4)
Bladder	1 (0.7)
Bone	1 (0.7)
Brain	1 (0.7)
Fallopian tube	1 (0.7)
Lymphedema	1 (0.7)
Mesenteric	1 (0.7)
Sarcoma	1 (0.7)
Synovial	1 (0.7)
Testicular	1 (0.7)
Uterine	1 (0.7)
Miscellaneous or other	6 (4.1)

<sup>a</sup>Types of cancers (n=148). Individuals with cancer may have reported >1 cancer type. Segments of both patients and surrogate seekers did not provide cancer type.

**Table 3.** Sociodemographic characteristics (surrogate seeker sample, N=97).

Variables	Values
Age (years), mean (SD)	51.72 (19.99)
<b>Sex, n (%)</b>	
Male	17 (18)
Female	65 (67)
Missing	15 (15)
<b>Race, n (%)</b>	
White	69 (71)
Hispanic or Latino	3 (3)
Black or African American	4 (4)
Native American or American Indian	1 (1)
Asian or Pacific Islander	1 (1)
Other	4 (4)
Missing	15 (15)
<b>Education, n (%)</b>	
Less than high school	1 (1)
High school or General Education Development	6 (76)
Some college	13 (13)
Completed college	41 (42)
Completed some postgraduate	2 (2)
Master's degree	12 (12)
Other advanced degree beyond master's	3 (3)
Missing	19 (20)
<b>Marital status, n (%)</b>	
Single	31 (32)
Partnered	4 (4)
Married	47 (48)
Missing	15 (15)
<b>Geographic region, n (%)</b>	
Midwest	0 (0)
Northeast	1 (1)
Southeast	77 (79)
Southwest	3 (3)
West	0 (0)
Missing	16 (16)
<b>Type of cancer, n (%)<sup>a</sup></b>	
Breast	15 (18)
Skin (squamous cell carcinoma, basal cell carcinoma, melanoma, and Merkel cell)	9 (11)
Colon and rectal	9 (11)
Brain	7 (8)
Prostate	6 (7)
Blood (leukemia, lymphoma, and myeloma)	4 (3)
Lung	4 (3)

Variables	Values
Ovarian	4 (3)
Pancreatic	4 (3)
Bladder	3 (4)
Liver	2 (2)
Kidney	2 (2)
Oral	2 (2)
Uterine	2 (2)
Stomach	2 (2)
Head and neck	1 (1)
Endometrial	1 (1)
Esophageal	1 (1)
Appendix	1 (1)
Bone	1 (1)
Unknown (cannot remember, not sure yet, and unknown)	3 (4)

<sup>a</sup>Types of cancers (n=83). Surrogate seekers may have reported >1 cancer type. Segments of both patients and surrogate seekers did not provide the cancer type.

## eHealth Literacy

Tables 4 shows the scores for each eHealth literacy competency. On the basis of a 5-point Likert-type scale, participants reported an above-average eHealth literacy score across all 4 competencies. On average, participants “agreed” that they have the skills needed to successfully access and understand web-based health information (*functional eHealth literacy*) and

apply what they learned to their health situation (*translational eHealth literacy*). Participants reported neither agreeing nor disagreeing that they have the skills needed to successfully exchange (*communicative eHealth literacy*) and evaluate (*critical eHealth literacy*) web-based information about cancer. Pearson correlation coefficients report a strong, statistically significant association between scores from each eHealth literacy competency ( $r=0.54-0.70$ ;  $P<.001$ ).

**Table 4.** eHealth literacy competency scores.

Competency	Values, n	Values, mean (SD)	Values, median (range)
Functional eHealth literacy	244	4.06 (0.76)	4.00 (1-5)
Communicative eHealth literacy	239	3.29 (0.91)	3.20 (1-5)
Critical eHealth literacy	238	3.49 (0.75)	3.60 (1.60-5)
Translational eHealth literacy	239	3.98 (0.65)	4.00 (2-5)

## eHealth Literacy Differences Between Individuals With Cancer and Surrogate Seekers

Table 5 shows the results of 4 hierarchical linear regression models that examined the association of respondent type (surrogate seeker vs individual with cancer) with each eHealth literacy competency, adjusting for age, gender, race, education, and marital status. The models were statistically significant for each of the four eHealth literacies: (1) functional ( $F_{6,226}=4.91$ ,  $P<.001$ ;  $R^2=0.12$ ,  $R^{2adj}=0.09$ ); (2) communicative ( $F_{6,226}=4.47$ ,  $P<.001$ ;  $R^2=0.11$ ,  $R^{2adj}=0.08$ ); (3) critical ( $F_{6,225}=4.30$ ,  $P<.001$ ;  $R^2=0.10$ ,  $R^{2adj}=0.08$ ); and (4) translational ( $F_{6,226}=2.99$ ,  $P=.01$ ;  $R^2=0.07$ ,  $R^{2adj}=0.05$ ).

Functional ( $P=.51$ ), communicative ( $P=.31$ ), and critical eHealth literacy ( $P=.63$ ) scores did not statistically significantly differ between individuals with cancer and surrogate seekers after

adjusting for sociodemographic variables. The results of the final regression model demonstrated a trend between respondent status and translational eHealth literacy but was not statistically significant ( $\beta=-0.15$ , SE 0.09;  $P=.09$ ). This result, which should be interpreted with caution, suggests that individuals with cancer may have higher confidence in applying the information they find on the internet to their own health, than their surrogate seeker counterparts.

We also found statistically significant associations between sociodemographic variables and eHealth literacy competencies in step 1 of the hierarchical linear regression models, whether the participant was an individual with cancer or a surrogate seeker. Identifying as a female and reporting a college education resulted in a positive association with functional eHealth literacy ( $F_{5,227}=5.82$ ,  $P<.001$ ;  $R^2=.11$ ,  $R^{2adj}=0.09$ ). In the communicative eHealth literacy model, reporting a younger age and a college education was positively associated with confidence in



exchanging web-based health information, ( $F_{5,227}=5.15$ ,  $P<.001$ ;  $R^2=.10$ ,  $R^{2adj}=0.08$ ). Similarly, reporting a younger age and a college education and identifying as a female resulted in a positive association with critical eHealth literacy ( $F_{5,226}=5.13$ ,

$P<.001$ ;  $R^2=.10$ ,  $R^{2adj}=0.08$ ). Translational eHealth literacy, however, was only statistically significantly associated with having a college education ( $F_{5,227}=3.00$ ,  $P=.04$ ,  $R^2=.06$ ,  $R^{2adj}=0.04$ ).

**Table 5.** Regression of surrogate seeker versus patient status on eHealth literacy.

Variable	Functional, $\beta$ (SE; 95% CI)	Communicative, $\beta$ (SE; 95% CI)	Critical, $\beta$ (SE; 95% CI)	Translational, $\beta$ (SE; 95% CI)
<b>Step 1</b>				
Age (years)	-0.01 (0.01; -0.01 to 0.01)	-0.01 (0.01 <sup>a</sup> ; -0.02 to -0.01)	-0.01 (0.01 <sup>a</sup> ; -0.02 to -0.01)	.00 (0.00; -0.01 to 0.00)
Sex <sup>b</sup>	-0.34 (0.11 <sup>c</sup> ; -0.55 to -0.13)	-0.19 (0.13; -0.45 to 0.07)	-0.25 (0.11 <sup>d</sup> ; -0.46 to -0.04)	-0.10 (0.10; -0.29 to 0.09)
Race <sup>e</sup>	.05 (0.15; -0.25 to 0.35)	-0.11 (0.19; -0.48 to 0.25)	-0.11 (0.15; -0.40 to 0.19)	.19 (0.14; -0.07 to 0.46)
Education <sup>f</sup>	.41 (0.10 <sup>g</sup> ; 0.21 to 0.61)	.34 (0.13 <sup>g</sup> ; 0.10 to 0.59)	.27 (0.10 <sup>g</sup> ; 0.07 to 0.47)	.30 (0.10 <sup>g</sup> ; 0.12 to 0.48)
Marital status <sup>h</sup>	.17 (.10; -0.03 to 0.36)	-0.04 (0.12; -0.28 to 0.20)	.08 (0.10; -0.12 to 0.28)	-0.06 (0.09; -0.24 to 0.12)
<b>Step 2</b>				
Respondent <sup>i</sup>	.07 (0.10; -0.13 to 0.26)	.13 (0.12; -0.12 to 0.37)	-0.05 (0.10; -0.24 to 0.15)	-0.15 (0.09 <sup>f</sup> ; -0.33 to 0.03)

<sup>a</sup> $P<.01$ .

<sup>b</sup>Sex (1=male; 0=female).

<sup>c</sup> $P<.001$ .

<sup>d</sup> $P<.05$ .

<sup>e</sup>Race (1=White; 0=people of color).

<sup>f</sup>Education (1=college educated; 0=less than college educated).

<sup>g</sup> $P=.09$ .

<sup>h</sup>Marital status (1=married; 0=not married).

<sup>i</sup>Respondent (1=caregiver; 0=patient).

### Individuals With Cancer and Surrogate Seekers' Evaluations of Credibility

Tables 6 shows that approximately half of the individuals with cancer (70/169, 41%) and surrogate seekers (37/89, 42%) evaluated the credibility of web-based cancer information according to the channel from where it was disseminated. Chi-square analyses showed that the relationship between individuals with cancer and surrogate seekers using the channel of web-based cancer information to appraise its credibility was not statistically significant ( $P=.98$ ). Examples of common information channels reviewed by individuals with cancer include professional and academic institutions such as the Mayo and Cleveland clinics, the National Institutes of Health and

National Cancer Institute, medical universities and websites, as well as physicians, peer-reviewed journals, and "credible" websites with .edu or .gov URLs. Examples reviewed by surrogate seekers include university-based publications, health care clinicians, the Mayo Clinic and National Institutes of Health. A similar code, *determined source credibility*, identified that individuals with cancer (8/169, 4.7%) and surrogate seekers (3/89, 3%) looked up to see who the author of web-based health information was and if they themselves were deemed credible. One person said, "Made sure a medical professional wrote the article" (ID 127; White female, 68 years old, completed some college). These respondents also often looked to see if a health care clinician or physician was cited to determine whether the information provided was credible.

**Table 6.** Frequencies of codes reported by individuals with cancer (n=169) and surrogate seekers (n=89).

Code	Individual with cancer, n (%)	Surrogate seeker, n (%)
Determined channel credibility	70 (41.4)	37 (41.6)
Determined source credibility	8 (4.7)	3 (3.4)
Checked citations for scientific support	5 (2.9)	16 (18) <sup>a</sup>
Cross-referenced content with other web-based sources	52 (30.8) <sup>b</sup>	18 (20.2)
Cross-referenced content with recommendations from clinicians	6 (3.6)	3 (3.4)
Discussed content with clinician	11 (6.5)	5 (5.6)
Miscellaneous	13 (7.7)	6 (6.7)
Not coded	4 (2.4)	1 (1.1)

<sup>a</sup> $P < .001$ .<sup>b</sup> $P < .10$ .

One code that reflected a considerable difference between individuals with cancer and surrogate seekers was the mode of appraising credibility by *checking citations for scientific support* of the information. A greater proportion of surrogate seekers (16/89, 18%) than individuals with cancer (5/169, 2.9%) checked for citations in their search for scientific information, ( $\chi^2_{1, N=258}=17.6, P < .001$ ). Only 3% (5/169) of individuals with cancer checked the citations of web-based health information to determine whether the content provided was credible, whereas nearly 18% (16/89) of surrogate seekers checked citations and references. One surrogate seeker stated as follows:

*I checked the sources at the bottom of the website. If no sources (scholarly websites or government/organization website hosts) were provided, then I did not deem it credible.* [Surrogate seeker ID 265; White male, 23 years old, college graduate]

Other surrogate seekers said, “I decided if it was credible if I had ample sources and clear answers.” (surrogate seeker ID 275; Hispanic female, 21 years old, completed some college); “There were credible references that were less than five years old.” (surrogate seeker ID 344; Black female, 56 years old, college graduate).

The second most used strategy was *cross-referencing content with other web-based sources*. Chi-square analyses revealed a trend toward statistical significance between individuals with cancer and surrogate seekers, showing that individuals with cancer may be more likely to cross-check information with other web-based sources than surrogate seekers ( $P = .07$ ). Approximately 30.7% (52/169) of individuals with cancer cross-referenced materials with additional sources, compared with 20% (18/89) of surrogate seekers. An example of a surrogate seeker response includes, “I looked at the date it was written and compared with other websites to see if it was similar—I looked at WebMD and then looked at Mayo Clinic to confirm accuracy” (surrogate seeker ID 233; White female, 51 years old, college educated). Individuals with cancer expressed similar experiences with cross-referencing multiple websites, such as searching to see if information was repeated

on several websites and comparing information from similar web-based sources.

**Table 6** provides additional procedures used by individuals with cancer and surrogate seekers to evaluate the credibility of web-based health information; however, they were used less frequently and were not statistically significantly different between individuals with cancer and surrogate seekers. This included cross-referencing content with recommendations from health care clinicians (6/169, 3.6% individuals with cancer vs 3/89, 3% surrogate seekers;  $P = .94$ ); discussing content acquired on the web with a health care clinician was also used (11/169, 6.5% individuals with cancer vs 5/89, 6% surrogate seekers;  $P = .78$ ), in addition to determining source credibility to assist in the credibility appraisal of web-based cancer information (8/169, 4.7% individuals with cancer vs 3/89, 3% surrogate seekers;  $P = .61$ ). The representation for each of these credibility appraisal procedures was relatively small and requires further analysis to determine differences, if any, between individuals with cancer and surrogate seekers.

The action words *looked* (individuals with cancer 22/123, 17.9%; surrogate seekers 17/57, 30%), *compared* (individuals with cancer 17/123, 13.8%; surrogate seekers 8/57, 14%), *read or reading* (individuals with cancer 16/123, 13%; surrogate seekers 7/57, 12%), and *searched or researched or tried to find* (individuals with cancer 11/123, 8.9%; surrogate seekers 7/57, 12%) were used most often among this sample (**Table 7**). When segmented into the 4 eHealth literacies, surrogate seekers used functional eHealth literacy terminology approximately 56% (32/57) of the time, compared with 43.9% (54/123) of individuals with cancer. Individuals with cancer evaluated web-based health information using a critical eHealth literacy perspective 39.8% (49/123) of the time, whereas surrogate seekers used critical eHealth literacy skills approximately 32% (18/57) of the time. There was a wider range of functional eHealth literacy (individuals with cancer: 54/123, 43.9%; surrogate seekers: 32/57, 56%), action word utterances than critical (*IC*: 49/123, 39.8%; surrogate seekers: 23/57, 40%) communicative (individuals with cancer: 12/123, 9.8%; surrogate seekers: 5/57, 9%), and translational (individuals with cancer: 8/123, 6.5%; surrogate seekers: 2/57, 4%).

**Table 7.** Frequencies of action words reported by individuals with cancer (n=123) and surrogate seekers (n=57).

Action words	Patients, n (%)	Surrogate seekers, n (%)
<b>Functional eHealth literacy</b>		
Looked	22 (17.9)	17 (29.8)
Reviewed	3 (2.4)	0 (0)
Read or reading	16 (13)	7 (12.3)
Gathered	1 (0.8)	1 (1.8)
Texted	1 (0.8)	0 (0)
Searched or researched or tried to find	11 (8.9)	7 (12.3)
<b>Communicative eHealth literacy</b>		
Spoke or speaking or discussed	5 (4.1)	1 (1.8)
Asked	7 (5.7)	4 (7)
<b>Critical eHealth literacy</b>		
Checked or double-checked	9 (7.3)	4 (7)
Cross-referenced	5 (4.1)	2 (3.5)
Considered or thinking	6 (4.9)	0 (0)
Compared	17 (13.8)	8 (14)
Evaluated or screened	2 (1.6)	0 (0)
Confirmed or made sure or verified	4 (3.3)	1 (1.8)
Decided or deemed or noted	4 (3.3)	3 (5.3)
Assumed	2 (1.6)	0 (0)
<b>Translational eHealth literacy</b>		
Used or using	6 (4.9)	2 (3.5)
Tried	2 (1.6)	0 (0)

## Discussion

### Principal Findings

The purpose of this study was to evaluate the eHealth literacy of individuals with cancer and surrogate seekers and explore the unique processes each group uses to evaluate web-based cancer information. Functional, communicative, critical, and translational eHealth literacy scores did not statistically significantly differ between individuals with cancer and surrogate seekers; however, we found differences in how individuals with cancer and surrogate seekers determine whether web-based cancer information is credible. This brings into question the validity of the web-based content retrieved from each group and how it influences health decisions, behaviors, and outcomes. The results demonstrate the value of understanding both the skill set and the process by which web-based content is accessed and evaluated before its exchange with others.

We did not find any statistically significant differences between individuals with cancer and surrogate seekers' confidence in their abilities to access, exchange, evaluate, and act on web-based health information for the purposes of maintaining or improving health. Individuals with cancer and surrogate seekers reported a high degree of confidence in their eHealth literacy across all competencies; however, functional and

translational eHealth literacy had a slightly higher average score than critical and communicative eHealth literacy scores.

### Comparison With Prior Work

The findings from this study align with previous literature that showed younger, more educated populations having higher eHealth literacy scores and an increased ease of accessing web-based content, often influenced by the level of use [17,20,38]. As this divergence in groups becomes more apparent, future research may examine the various factors that inhibit the communicative and critical evaluation skills of older groups of individuals with cancer and surrogate seekers.

Heiman et al [39] found that the internet was the third most important source of information for individuals with cancer, preceded only by their oncologist and print media. They also discovered that the biggest concern for individuals with cancer was not being able to differentiate between reliable and unreliable websites when searching for information pertaining to their diagnosis and treatment [39]. Our study assessed exactly how individuals with cancer and surrogate seekers evaluated the credibility of web-based health information, which provides further knowledge and understanding to expand upon previous research.

Regardless of their eHealth literacy skill sets, individuals with cancer and surrogate seekers most often determine credibility

according to the channel of the web-based cancer information. This suggests that those seeking cancer information do not see a significant need for corroboration of web-based information if the site publishing it is perceived as credible. Several respondents noted that they determined whether web-based health information was credible based on whether they had visited a website in the past and were already familiar with it. Others determined that the information was credible if the website provided the information that the patient or surrogate seeker was searching for. Similar findings have been discussed in the scoping review by Verm et al [40], which found that eHealth literacy was positively correlated with surrogate seekers' strategies enacted such as seeking a second opinion, awareness of treatment options, shared decision-making, and trust in the health care system. An important factor to consider, moving forward, when analyzing how people appraise and evaluate the credibility of web-based health information includes several cognitive biases such as confirmation bias.

Confirmation bias, or the phenomena of "seeking or interpreting evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand" has the potential to greatly influence the subconscious motivations for seeking web-based health information in patients and supporting network roles [41]. Meppelink et al [42] recently found that individuals with high health literacy in web-based health information seeking tend to select belief-consistent information that is rated as credible, useful, and convincing. People who are prone to confirmation bias are considered to be overconfident in their knowledge and skills in evaluating content [43]. Therefore, purposefully seeking information that confirms and validates prior assumptions could negatively impact a patient's or surrogate seekers' understanding of a diagnosis, treatment, and health care management and have detrimental impacts on advocacy skills throughout a health care experience. Future research is needed to examine whether this biased information search is fueled by a lack of knowledge acquisition skills or a degree of managing uncertainty surrounding a cancer diagnosis of themselves or their loved ones.

More surrogate seekers than individuals with cancer reviewed the references, citations, and links provided with web-based health information to determine whether the content was credible. Conversely, more individuals with cancer than surrogate seekers cross-referenced content with other web-based sources to determine its credibility. Individuals with cancer may determine the credibility of information through sheer quantity (ie, how often it is repeated across multiple sources), whereas surrogate seekers may determine its credibility based on scientific support and quality of citations. This distinction in information-seeking behaviors and preferred evaluation methods between individuals with cancer and surrogate seekers is important to examine, as individuals with cancer are prone to misinformation and to consensus effects of information [44]. Simply accessing and confirming that a piece of information is available from more than one location does not guarantee its credibility, and dilemmas such as this could inadvertently perpetuate the spread and use of misinformation pertaining to cancer care. In-depth qualitative inquiry is needed to examine the tendencies of individuals with cancer to acquire information

(ie, referring to multiple sources and researching patient experiences) compared with surrogate seekers' preferences for a more scientifically grounded knowledge base (ie, information channels and confirmation of acquired information).

Identifying the diverse and unique ways in which individuals with cancer and surrogate seekers access, appraise, and evaluate the credibility of web-based cancer information could provide a deeper, more tailored design and evaluation of patient education resources. These resources are developed with the intention of being perceived as credible by each recipient, so having distinctive information on how different groups retrieve this web-based content could help us better appeal to the established behaviors that individuals with cancer and surrogate seekers use to enhance their appraisal and evaluation of the credibility of web-based information. Understanding the process by which individuals with cancer and their surrogate seekers access and evaluate web-based information credibility will further inform how to deliver educational content from varying web-based sources, hopefully increasing both patient and surrogate seekers' autonomy and self-efficacy throughout their cancer care. While identifying the dynamic ways in which individuals with cancer and surrogate seekers access and appraise web-based cancer information yields important insights into future message design and implementation, understanding the interpersonal contexts of this population is imperative for a more refined understanding of why they execute such skills. Researchers should consider nuances related to psychological and relational factors that affect these appraisal and evaluation skills more deeply, including the level of perceived importance that the patient or surrogate seeker has for receiving credible web-based information, how their personal relationships impact their appraisal and evaluation skills, and how stress levels impact the appraisal and evaluation process.

### Limitations

The limitations of this study include addressing the longitudinal effects of eHealth literacy, assessing contextual factors related to individuals with blood cancer, and examining dyadic groups of individuals with cancer and surrogate seekers. First, this was a cross-sectional study, which poses a challenge given that eHealth literacy is a dynamic skill set that evolves over time, making it difficult to determine large effects from 1 period. Future surveillance research is needed to explore how eHealth literacy in individuals with cancer and surrogate seeker groups changes over time. This study explored eHealth literacy in 2 independent groups of individuals with cancer and surrogate seekers. The TeHLI is a relatively new eHealth literacy measure. Similar to the eHEALS, we recognize that more advanced statistical analyses must be conducted to strengthen evidence for its use (eg, measurement invariance and item response theory). Examining measurement invariance in future studies is particularly important as this statistical test is the only way to confirm whether a latent variable can be truly compared across 2 or more groups. Given the exploratory nature of this study, such tests were not conducted.

This study did not take possible contextual factors for individuals with cancer and surrogate seekers into account, such as date of diagnosis, how recently they received their diagnosis

compared with when they searched for web-based health information, or the varying levels of stress experienced when participating in this web-based health information seeking and credibility appraisal process. Assessing these factors in future studies could assist in better understanding the unique circumstances experienced within the population and how these interpersonal factors possibly influence eHealth literacy over the course of their cancer journey.

Most respondents in this study were diagnosed with or had cared for someone with breast, skin, or some type of blood cancer. Although not consistent with national estimates of cancer incidence and prevalence [45], the results are consistent with the estimates reported in the catchment area from which these data were collected [46]. Future research should take a more stratified approach to recruitment for cancer type to ensure that region-specific nuances are identified and controlled. In addition, replicating this study among a sufficient sample of individuals with cancer and surrogate seekers dealing with a specific cancer through subgroup analyses will be important for deepening the understanding of any possible differences in eHealth literacy between individuals with cancer and surrogate seekers. The cumulative sample size for this study was relatively small and was restricted to the southeastern region of the nation. The results were limited to a specific region of the United States, and their generalization should be approached with caution. Future research would also benefit from observing individuals with cancer and surrogate seekers as they evaluate web-based cancer information, rather than relying on self-reported procedures that are prone to reporting biases. Regardless of these limitations, the data were derived from validated, theoretically-driven measures and we used a mixed methods approach to achieve the purpose of this study.

### Practical and Theoretical Implications

Using the TMeHL as a theoretical foundation [17], this study aimed to discern how individuals with cancer and surrogate seekers evaluated the credibility of web-based health information. The TMeHL provided an established intrapersonal skill set pertaining to eHealth literacy, which highlighted several discrete skills that were used to interpret the results. From this, we gathered information that showed that eHealth literacy scores remained similar between individuals with cancer and surrogate seekers. However, these groups use their skills in unique and diverse ways. For example, they weigh the features of web-based health information differently when they appraise and evaluate credibility. These differences include source, channel, scientific references, quantity of web-based sources viewed, etc. The features of critical eHealth literacy were centralized around individuals with cancer and surrogate seekers' descriptions of how they evaluate web-based cancer information. However, skills relevant to all 4 eHealth literacies represented in the TMeHL were represented across these descriptions, suggesting that these skills function together.

Seeing the variance between individuals with cancer and surrogate seekers in what they deem most important when evaluating the credibility of web-based health information offers several avenues for future research, including exploring the potential barriers each group encounters when searching for and appraising web-based information. According to the TMeHL [17], an individual's eHealth literacy skill set may evolve to proactively overcome challenges that are persistent in their environment and pose threats to their ability to effectively navigate web-based health information. It may be that individuals with cancer or surrogate seekers may each experience different challenges than the other that must be addressed to properly ascertain the appraisal processes for each group. Psychological and relational considerations may also be incorporated to better distinguish between these groups, such as level of support network load or burden, perceived burden of patients with cancer, and varying degrees of self-efficacy and motivation needed to advocate for the importance of credible web-based cancer information related to their health. Further research should also focus on the clear preferences that individuals with cancer and surrogate seekers have that revolve around using sources versus channels to establish whether information is credible.

When viewed practically, strategic messages that are targeted to a group's preferred source or channel of information will increase the likelihood of its seeing the information as relevant and credible and have greater potential to better initiate patient and surrogate seeker engagement over the course of one's own or a loved one's health care management. The continual advances in computer-assisted technologies make tailoring messages to these variables an important next step, as this level of personalization can not only help enhance the individuals with cancer and surrogate seekers' web-based health information acquisition experience, but also hold considerable potential to provide pertinent information to these populations. Our study included patients and surrogate seekers, but we did not examine the eHealth literacy skills or information-seeking behaviors of individuals with cancer with their own surrogate seekers. Future research with dyads of individuals with cancer and health care clinicians is needed to determine the value of tailored messages within this context.

### Conclusions

Individuals with cancer and surrogate seekers report similar eHealth literacy levels, but there is evidence that these groups apply unique approaches to evaluating the credibility of web-based health information. The results of this study have important theoretical and practical implications for expanding the understanding and applicability of the TMeHL to inform future message design interventions. Future research is needed to examine how dyads of individuals with cancer and surrogate seekers evaluate web-based health information and the acceptability of collaborative patient with cancer support network dyadic eHealth literacy interventions.

## Acknowledgments

This work was supported by the University of Florida Clinical and Translational Science Institute grant support (National Institutes of Health National Center for Advancing Translational Sciences under grant UL1 TR000064) and the University of Florida College of Journalism and Communications Dean's Seed Grant Funding.

## Data Availability

The data set from which the results of this study were derived can be obtained from the corresponding author upon request.

## Conflicts of Interest

SRP is an employee of Johnson & Johnson.

### Multimedia Appendix 1

Sociodemographic questions from Health Information National Trends Survey and US Census Bureau.

[[DOCX File , 12 KB - formative\\_v6i9e36714\\_app1.docx](#) ]

### Multimedia Appendix 2

Surrogate seeker types.

[[DOCX File , 13 KB - formative\\_v6i9e36714\\_app2.docx](#) ]

## References

1. Weir HK, Thompson TD, Stewart SL, White MC. Cancer incidence projections in the United States between 2015 and 2050. *Prev Chronic Dis* 2021 Jun 10;18:E59 [[FREE Full text](#)] [doi: [10.5888/pcd18.210006](https://doi.org/10.5888/pcd18.210006)] [Medline: [34114543](https://pubmed.ncbi.nlm.nih.gov/34114543/)]
2. van Ryn M, Sanders S, Kahn K, van Houtven C, Griffin JM, Martin M, et al. Objective burden, resources, and other stressors among informal cancer caregivers: a hidden quality issue? *Psychooncology* 2011 Jan;20(1):44-52 [[FREE Full text](#)] [doi: [10.1002/pon.1703](https://doi.org/10.1002/pon.1703)] [Medline: [20201115](https://pubmed.ncbi.nlm.nih.gov/20201115/)]
3. Kent EE, Rowland JH, Northouse L, Litzelman K, Chou WY, Shelburne N, et al. Caring for caregivers and patients: research and clinical priorities for informal cancer caregiving. *Cancer* 2016 Jul 01;122(13):1987-1995 [[FREE Full text](#)] [doi: [10.1002/cncr.29939](https://doi.org/10.1002/cncr.29939)] [Medline: [26991807](https://pubmed.ncbi.nlm.nih.gov/26991807/)]
4. Vahdat S, Hamzehgardeshi L, Hessam S, Hamzehgardeshi Z. Patient involvement in health care decision making: a review. *Iran Red Crescent Med J* 2014 Jan;16(1):e12454 [[FREE Full text](#)] [doi: [10.5812/ircmj.12454](https://doi.org/10.5812/ircmj.12454)] [Medline: [24719703](https://pubmed.ncbi.nlm.nih.gov/24719703/)]
5. Lam TC, Balboni TA, Maciejewski PK, Prigerson HG. Differences between patient and caregiver ratings of advanced cancer patients' quality of life and correlates of their disagreement. *J Clin Oncol* 2014 May 20;32(15\_suppl):9560. [doi: [10.1200/jco.2014.32.15\\_suppl.9560](https://doi.org/10.1200/jco.2014.32.15_suppl.9560)]
6. Glajchen M. The emerging role and needs of family caregivers in cancer care. *J Support Oncol* 2004;2(2):145-155. [Medline: [15328817](https://pubmed.ncbi.nlm.nih.gov/15328817/)]
7. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. *Med Decis Making* 2015 Jan;35(1):114-131 [[FREE Full text](#)] [doi: [10.1177/0272989X14551638](https://doi.org/10.1177/0272989X14551638)] [Medline: [25351843](https://pubmed.ncbi.nlm.nih.gov/25351843/)]
8. Sainio C, Eriksson E, Lauri S. Patient participation in decision making about care. *Cancer Nurs* 2001 Jun;24(3):172-179. [doi: [10.1097/00002820-200106000-00002](https://doi.org/10.1097/00002820-200106000-00002)] [Medline: [11409060](https://pubmed.ncbi.nlm.nih.gov/11409060/)]
9. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. *JAMA* 2001 May 23;285(20):2612-2621 [[FREE Full text](#)] [doi: [10.1001/jama.285.20.2612](https://doi.org/10.1001/jama.285.20.2612)] [Medline: [11368735](https://pubmed.ncbi.nlm.nih.gov/11368735/)]
10. Maddock C, Lewis I, Kafait A. 3643 POSTER European patient online information needs. *Eur J Cancer* 2011 Sep;47:S271-S272. [doi: [10.1016/s0959-8049\(11\)71240-0](https://doi.org/10.1016/s0959-8049(11)71240-0)]
11. Cutrona SL, Mazor KM, Vieux SN, Luger TM, Volkman JE, Finney Rutten LJ. Health information-seeking on behalf of others: characteristics of "surrogate seekers". *J Cancer Educ* 2015 Mar;30(1):12-19 [[FREE Full text](#)] [doi: [10.1007/s13187-014-0701-3](https://doi.org/10.1007/s13187-014-0701-3)] [Medline: [24989816](https://pubmed.ncbi.nlm.nih.gov/24989816/)]
12. Viswanath K. Science and society: the communications revolution and cancer control. *Nat Rev Cancer* 2005 Oct;5(10):828-835. [doi: [10.1038/nrc1718](https://doi.org/10.1038/nrc1718)] [Medline: [16195753](https://pubmed.ncbi.nlm.nih.gov/16195753/)]
13. Abrahamson JA, Fisher KE. "What's past is prologue": towards a general model of lay information mediary behaviour. *Inf Res* 2007 Oct;12(4):15 [[FREE Full text](#)]
14. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. *Arch Intern Med* 2005 Dec 12;165(22):2618-2624. [doi: [10.1001/archinte.165.22.2618](https://doi.org/10.1001/archinte.165.22.2618)] [Medline: [16344419](https://pubmed.ncbi.nlm.nih.gov/16344419/)]
15. Peterson JS, Swire-Thompson B, Johnson SB. What is the alternative? Responding strategically to cancer misinformation. *Future Oncol* 2020 Sep;16(25):1883-1888. [doi: [10.2217/fon-2020-0440](https://doi.org/10.2217/fon-2020-0440)] [Medline: [32564627](https://pubmed.ncbi.nlm.nih.gov/32564627/)]

16. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. *J Med Internet Res* 2006 Jun 16;8(2):e9 [FREE Full text] [doi: [10.2196/jmir.8.2.e9](https://doi.org/10.2196/jmir.8.2.e9)] [Medline: [16867972](https://pubmed.ncbi.nlm.nih.gov/16867972/)]
17. Paige SR, Stellefson M, Krieger JL, Miller MD, Cheong J, Anderson-Lewis C. Transactional eHealth literacy: developing and testing a multi-dimensional instrument. *J Health Commun* 2019;24(10):737-748 [FREE Full text] [doi: [10.1080/10810730.2019.1666940](https://doi.org/10.1080/10810730.2019.1666940)] [Medline: [31583963](https://pubmed.ncbi.nlm.nih.gov/31583963/)]
18. Levin-Zamir D, Bertschi I. Media health literacy, eHealth literacy, and the role of the social environment in context. *Int J Environ Res Public Health* 2018 Aug 03;15(8):1643 [FREE Full text] [doi: [10.3390/ijerph15081643](https://doi.org/10.3390/ijerph15081643)] [Medline: [30081465](https://pubmed.ncbi.nlm.nih.gov/30081465/)]
19. Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C. Proposing a transactional model of eHealth literacy: concept analysis. *J Med Internet Res* 2018 Oct 02;20(10):e10175 [FREE Full text] [doi: [10.2196/10175](https://doi.org/10.2196/10175)] [Medline: [30279155](https://pubmed.ncbi.nlm.nih.gov/30279155/)]
20. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. *J Med Internet Res* 2015 Mar 17;17(3):e70 [FREE Full text] [doi: [10.2196/jmir.3992](https://doi.org/10.2196/jmir.3992)] [Medline: [25783036](https://pubmed.ncbi.nlm.nih.gov/25783036/)]
21. Yuen EY, Knight T, Ricciardelli LA, Burney S. Health literacy of caregivers of adult care recipients: a systematic scoping review. *Health Soc Care Community* 2018 Mar;26(2):e191-e206. [doi: [10.1111/hsc.12368](https://doi.org/10.1111/hsc.12368)] [Medline: [27426731](https://pubmed.ncbi.nlm.nih.gov/27426731/)]
22. Nijboer C, Triemstra M, Tempelaar R, Mulder M, Sanderman R, van den Bos GA. Patterns of caregiver experiences among partners of cancer patients. *Gerontologist* 2000 Dec;40(6):738-746. [doi: [10.1093/geront/40.6.738](https://doi.org/10.1093/geront/40.6.738)] [Medline: [11131090](https://pubmed.ncbi.nlm.nih.gov/11131090/)]
23. 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](https://doi.org/10.1016/j.socscimed.2007.01.012)] [Medline: [17328998](https://pubmed.ncbi.nlm.nih.gov/17328998/)]
24. How online family caregivers interact with their health online. Pew Research Center. 2013 Jun 20. URL: <https://www.pewresearch.org/internet/2013/06/20/how-online-family-caregivers-interact-with-their-health-online/> [accessed 2021-01-18]
25. Wright KB, Rains SA. Weak tie support preference and preferred coping styles as predictors of perceived credibility within health-related computer-mediated support groups. *Health Commun* 2014;29(3):281-287. [doi: [10.1080/10410236.2012.751084](https://doi.org/10.1080/10410236.2012.751084)] [Medline: [23790220](https://pubmed.ncbi.nlm.nih.gov/23790220/)]
26. Sbaifi L, Rowley J. Trust and credibility in Web-based health information: a review and agenda for future research. *J Med Internet Res* 2017 Jun 19;19(6):e218 [FREE Full text] [doi: [10.2196/jmir.7579](https://doi.org/10.2196/jmir.7579)] [Medline: [28630033](https://pubmed.ncbi.nlm.nih.gov/28630033/)]
27. Thompson TL. Online health information credibility. In: Thompson TL, editor. *Encyclopedia of Health Communication*. Thousand Oaks, CA, USA: Sage Publications; 2014.
28. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. *J Med Internet Res* 2019 May 02;21(5):e12522 [FREE Full text] [doi: [10.2196/12522](https://doi.org/10.2196/12522)] [Medline: [31045507](https://pubmed.ncbi.nlm.nih.gov/31045507/)]
29. Neter E, Brainin E. Perceived and performed eHealth literacy: survey and simulated performance test. *JMIR Hum Factors* 2017 Jan 17;4(1):e2 [FREE Full text] [doi: [10.2196/humanfactors.6523](https://doi.org/10.2196/humanfactors.6523)] [Medline: [28096068](https://pubmed.ncbi.nlm.nih.gov/28096068/)]
30. 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](https://doi.org/10.1016/j.jbi.2008.08.010)] [Medline: [18929686](https://pubmed.ncbi.nlm.nih.gov/18929686/)]
31. Griebel L, Enwald H, Gilstad H, Pohl AL, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? *Inform Health Soc Care* 2018 Dec;43(4):427-442. [doi: [10.1080/17538157.2017.1364247](https://doi.org/10.1080/17538157.2017.1364247)] [Medline: [29045164](https://pubmed.ncbi.nlm.nih.gov/29045164/)]
32. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. *J Med Internet Res* 2011 Dec 23;13(4):e125 [FREE Full text] [doi: [10.2196/jmir.2035](https://doi.org/10.2196/jmir.2035)] [Medline: [22193243](https://pubmed.ncbi.nlm.nih.gov/22193243/)]
33. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnett S. What is eHealth (6)? Development of a conceptual model for eHealth: qualitative study with key informants. *J Med Internet Res* 2017 Oct 24;19(10):e324 [FREE Full text] [doi: [10.2196/jmir.8106](https://doi.org/10.2196/jmir.8106)] [Medline: [29066429](https://pubmed.ncbi.nlm.nih.gov/29066429/)]
34. Barnlund DC. A transactional model of communication. In: Sereno KK, Mortensen CD, editors. *Foundations of Communication Theory*. Manhattan, NY, USA: Harper and Row; 1970:83-102.
35. Creswell JW, Plano Clark VL. *Designing and Conducting Mixed Methods Research*. 3rd edition. Thousand Oaks, CA, USA: Sage Publications; 2018.
36. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005 Nov;15(9):1277-1288. [doi: [10.1177/1049732305276687](https://doi.org/10.1177/1049732305276687)] [Medline: [16204405](https://pubmed.ncbi.nlm.nih.gov/16204405/)]
37. Morse JM. Approaches to qualitative-quantitative methodological triangulation. *Nurs Res* 1991;40(2):120-123. [doi: [10.1097/00006199-199103000-00014](https://doi.org/10.1097/00006199-199103000-00014)] [Medline: [2003072](https://pubmed.ncbi.nlm.nih.gov/2003072/)]
38. Battineni G, Baldoni S, Chintalapudi N, Sagaro GG, Pallotta G, Nittari G, et al. Factors affecting the quality and reliability of online health information. *Digit Health* 2020 Aug 30;6:2055207620948996 [FREE Full text] [doi: [10.1177/2055207620948996](https://doi.org/10.1177/2055207620948996)] [Medline: [32944269](https://pubmed.ncbi.nlm.nih.gov/32944269/)]
39. Heiman H, Keinki C, Huebner J, Working Group Prevention and Integrative Oncology of the German Cancer Society. EHealth literacy in patients with cancer and their usage of Web-based information. *J Cancer Res Clin Oncol* 2018 Sep;144(9):1843-1850. [doi: [10.1007/s00432-018-2703-8](https://doi.org/10.1007/s00432-018-2703-8)] [Medline: [30006768](https://pubmed.ncbi.nlm.nih.gov/30006768/)]
40. Verma R, Saldanha C, Ellis U, Sattar S, Haase KR. eHealth literacy among older adults living with cancer and their caregivers: a scoping review. *J Geriatr Oncol* 2022 Jun;13(5):555-562. [doi: [10.1016/j.jgo.2021.11.008](https://doi.org/10.1016/j.jgo.2021.11.008)] [Medline: [34810146](https://pubmed.ncbi.nlm.nih.gov/34810146/)]

41. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol* 1998 Jun 01;2(2):175-220. [doi: [10.1037/1089-2680.2.2.175](https://doi.org/10.1037/1089-2680.2.2.175)]
42. Meppelink CS, Smit EG, Franssen ML, Diviani N. "I was right about vaccination": confirmation bias and health literacy in online health information seeking. *J Health Commun* 2019;24(2):129-140. [doi: [10.1080/10810730.2019.1583701](https://doi.org/10.1080/10810730.2019.1583701)] [Medline: [30895889](https://pubmed.ncbi.nlm.nih.gov/30895889/)]
43. Kenski K. Overcoming confirmation and blind spot biases when communicating science. In: Jamieson KH, Kahan DM, Scheufele DA, editors. *The Oxford Handbook of the Science of Science Communication*. Oxford, UK: Oxford University Press; 2017:369-376.
44. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. *Annu Rev Public Health* 2020 Apr 02;41:433-451. [doi: [10.1146/annurev-publhealth-040119-094127](https://doi.org/10.1146/annurev-publhealth-040119-094127)] [Medline: [31874069](https://pubmed.ncbi.nlm.nih.gov/31874069/)]
45. Cancer Stat Facts: Common Cancer Sites. National Cancer Institute. 2020. URL: <https://seer.cancer.gov/statfacts/html/common.html> [accessed 2021-01-18]
46. Paige SR, Alpert JM, Bylund CL. Fatalistic cancer beliefs across generations and geographic classifications: examining the role of health information seeking challenges and confidence. *J Cancer Educ* 2021 Feb;36(1):3-9 [FREE Full text] [doi: [10.1007/s13187-020-01820-3](https://doi.org/10.1007/s13187-020-01820-3)] [Medline: [32648238](https://pubmed.ncbi.nlm.nih.gov/32648238/)]

## Abbreviations

**eHEALS:** eHealth Literacy Scale

**REDCap:** Research Electronic Data Capture

**TeHLI:** Transactional eHealth Literacy Instrument

**TMeHL:** Transactional Model of eHealth Literacy

*Edited by A Mavragani; submitted 21.01.22; peer-reviewed by L Sudbury-Riley, M Zaidi, A Efthymiou; comments to author 11.04.22; revised version received 24.05.22; accepted 15.08.22; published 28.09.22.*

*Please cite as:*

*Vasquez TS, Bylund CL, Alpert J, Close J, Le T, Markham MJ, Taylor GB, Paige SR*

*Comparing Transactional eHealth Literacy of Individuals With Cancer and Surrogate Information Seekers: Mixed Methods Study*  
*JMIR Form Res* 2022;6(9):e36714

URL: <https://formative.jmir.org/2022/9/e36714>

doi: [10.2196/36714](https://doi.org/10.2196/36714)

PMID: [36170007](https://pubmed.ncbi.nlm.nih.gov/36170007/)

©Taylor S Vasquez, Carma L Bylund, Jordan Alpert, Julia Close, Tien Le, Merry Jennifer Markham, Greenberry B Taylor, Samantha R Paige. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 28.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Usability Testing of the Kidney Score Platform to Enhance Communication About Kidney Disease in Primary Care Settings: Qualitative Think-Aloud Study

Delphine S Tuot<sup>1</sup>, MD; Susan T Crowley<sup>2</sup>, MD; Lois A Katz<sup>3</sup>, MD; Joseph Leung<sup>3</sup>, MD; Delly K Alcantara-Cadillo<sup>4</sup>, MD; Christopher Ruser<sup>2</sup>, MD; Elizabeth Talbot-Montgomery<sup>5</sup>, BS; Joseph A Vassalotti<sup>6</sup>, MD

<sup>1</sup>University of California San Francisco, San Francisco, CA, United States

<sup>2</sup>Veterans Affairs Connecticut Healthcare System, West Haven, CT, United States

<sup>3</sup>Veterans Affairs New York Harbor Healthcare System, New York, NY, United States

<sup>4</sup>Yale University, New Haven, CT, United States

<sup>5</sup>National Kidney Foundation, New York, NY, United States

<sup>6</sup>Icahn School of Medicine at Mount Sinai, New York, NY, United States

**Corresponding Author:**

Delphine S Tuot, MD

University of California San Francisco

1001 Potrero Ave

San Francisco, CA, 94110

United States

Phone: 1 628 206 3784

Email: [Delphine.tuot@ucsf.edu](mailto:Delphine.tuot@ucsf.edu)

## Abstract

**Background:** Patient awareness of chronic kidney disease (CKD) is low in part due to suboptimal testing for CKD among those at risk and lack of discussions about kidney disease between patients and clinicians. To bridge these gaps, the National Kidney Foundation developed the Kidney Score Platform, which is a web-based series of tools that includes resources for health care professionals as well as an interactive, dynamic patient-facing component that includes a brief questionnaire about risk factors for kidney disease, individualized assessment of risk for developing CKD, and self-management tools to manage one's kidney disease.

**Objective:** The aim of this study is to perform usability testing of the patient component of the Kidney Score platform among veterans with and at risk for kidney disease and among clinicians working as primary care providers in Veterans Affairs administration.

**Methods:** Think-aloud exercises were conducted, during which participants (veterans and clinicians) engaged with the platform while verbalizing their thoughts and making their perceptions, reasonings, and decision points explicit. A usability facilitator observed participants' behaviors and probed selectively to clarify their comprehension of the tool's instructions, content, and overall functionality. Thematic analysis on the audio-recording transcripts was performed, focusing on positive attributes, negative comments, and areas that required facilitator involvement.

**Results:** Veterans (N=18) were 78% (14/18) male with a mean age of 58.1 years. Two-thirds (12/18) were of non-White race/ethnicity, 28% (5/18) had laboratory evidence of CKD without a formal diagnosis, and 50% (9/18) carried a diagnosis of hypertension or diabetes. Clinicians (N=19) were 29% (5/17) male, 30% (5/17) of non-White race/ethnicity, and had a mean of 17 (range 4-32) years of experience. Veterans and clinicians easily navigated the online tool and appreciated the personalized results page as well as the inclusion of infographics to deliver key educational messages. Three major themes related to content and communication about risk for CKD emerged from the think-aloud exercises: (1) tension between lay and medical terminology when discussing kidney disease and diagnostic tests, (2) importance of linking general information to concrete self-management actions, and (3) usefulness of the tool as an adjunct to the office visit to prepare for patient-clinician communication. Importantly, these themes were consistent among interviews involving both veterans and clinicians.

**Conclusions:** Veterans and clinicians both thought that the Kidney Score Platform would successfully promote communication and discussion about kidney disease in primary care settings. Tension between using medical terminology that is used regularly

by clinicians versus lay terminology to promote CKD awareness was a key challenge, and knowledge of this can inform the development of future CKD educational materials.

(*JMIR Form Res* 2022;6(9):e40001) doi:[10.2196/40001](https://doi.org/10.2196/40001)

## KEYWORDS

chronic kidney disease; CKD; awareness; usability; kidney; renal; think aloud; self-management; patient education; health education

## Introduction

Chronic kidney disease (CKD) is a chronic disease that requires individual participation in health-related behaviors to decrease the risk of progression and associated cardiovascular disease [1]. Patient awareness of CKD—including the knowledge of having a kidney problem, the perceived risk of developing kidney disease, and ability to affect one's kidney health—is necessary for patients to participate in shared decision-making about their kidney health and to apply management recommendations to improve outcomes [2]. However, as many as half of patients with advanced CKD are unaware that they have kidney disease, including those at high risk for kidney function decline [3] and those with laboratory manifestations of their kidney disease [4,5].

Reasons for the low prevalence of CKD awareness among individuals with CKD are varied and include patient, provider, and health system factors [6]. Two of these contributing factors are suboptimal testing for CKD among those at risk for kidney disease [7,8] and lack of discussions about kidney disease between patients and clinicians among those individuals with laboratory-documented CKD. Studies in primary care settings have consistently demonstrated that discussions about kidney disease occur less frequently than do conversations about other chronic diseases [9] and that primary care clinicians experience challenges in improving their patients' understanding of kidney disease, even when using principles of shared decision-making [10,11]. Additionally, individuals at risk for CKD have low perceived risk of the condition [12], which may exacerbate primary care clinicians' concerns of emotionally overwhelming patients with a diagnosis of CKD [6].

To bridge the communication gap about kidney disease among patients and health care professionals and increase testing among individuals at risk for CKD, the National Kidney Foundation (NKF) developed the Kidney Score Platform, leveraging the

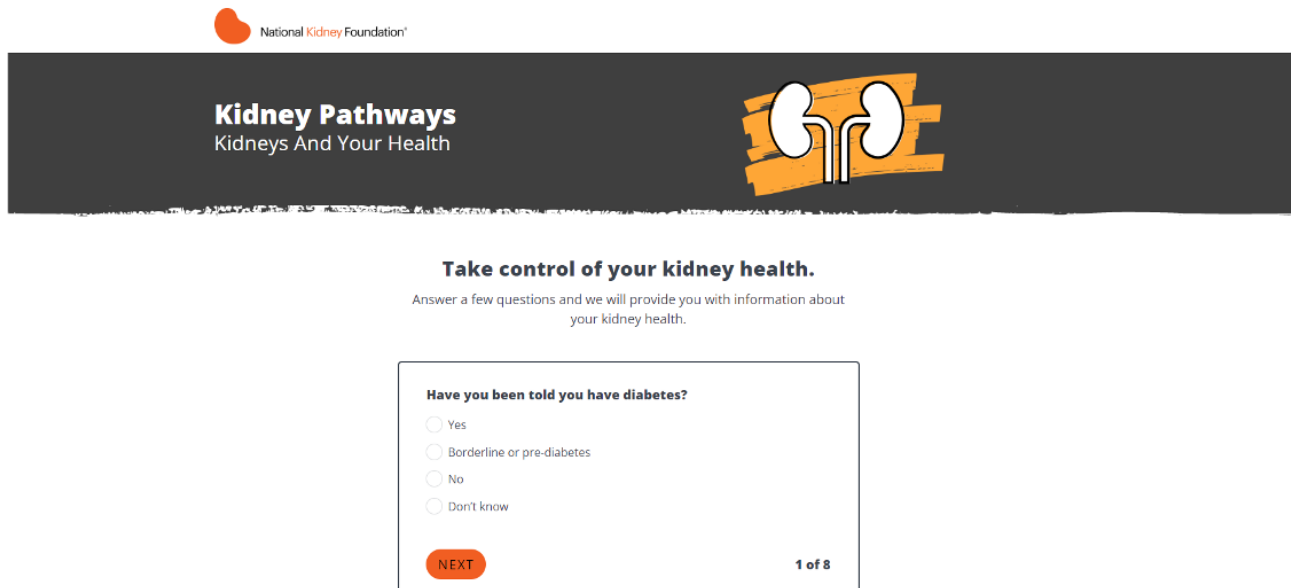
behavior change wheel, a validated framework used to design interventions to incite individual behavior change [13]. The Kidney Score Platform is a web-based series of tools that includes resources for health care professionals to encourage the use of a population health strategy for CKD management [14] as well as an interactive, dynamic component for patients to increase their knowledge and self-management. Before embarking on a study of the Kidney Score Platform's impact on patient-clinician communication about kidney disease and individual awareness of CKD [13], we sought feedback from clinicians about the provider resources available on the Kidney Score Platform and engaged in usability testing of the patient-facing tool to gather feedback regarding its acceptability and potential use [15]. A partnership between the NKF and Veterans Administration provided an opportunity to perform usability testing of the patient component of the Kidney Score Platform among veterans with and at risk for kidney disease and among clinicians working as primary care providers in Veterans Affairs administration. Here, we describe our experience with that usability testing, which culminated in important refinements to the Kidney Score Platform, which is now ready for an examination of its impact on patient-clinician conversations about kidney disease in Veterans Administration ambulatory settings.

## Methods

### Patient Education Tool

The patient-facing component of the Kidney Score platform includes a brief questionnaire about risk factors for kidney disease that results in a personalized educational results page providing an individualized assessment of risk for developing CKD or self-management tools to manage one's kidney disease (Figures 1-3). Development of the tool has been described in depth elsewhere [13].

**Figure 1.** Example questions within the CKD risk self-assessment tool. CKD: chronic kidney disease.



National Kidney Foundation<sup>®</sup>

## Kidney Pathways

Kidneys And Your Health

### Take control of your kidney health.

Answer a few questions and we will provide you with information about your kidney health.

**Have you been told you have diabetes?**

Yes


Borderline or pre-diabetes

No

Don't know

**NEXT** **1 of 8**

**Figure 2.** Example self-assessment results, linking risk factors to kidney disease risk, providing education about CKD diagnostic tests, and encouraging patients to review diagnostic tests with their primary care clinician to increase awareness of CKD. CKD: chronic kidney disease.



[MENU](#)
[DONATE](#)


Your submission has been saved.

## You are at risk for Chronic Kidney Disease

### Understanding your results


The information you entered shows **you have 3 common risk factors** for chronic kidney disease (CKD). This means you have an increased chance of developing CKD. Because CKD has no early symptoms, people at risk for CKD should be tested regularly to find out about their kidney health.

The information below explains your risk factors. **Print a copy to discuss with your healthcare professional and ask them if CKD testing is right for you.**




#### Diabetes

Diabetes is a major risk factor for kidney disease. Diabetes can be treated and can improve with healthy lifestyle choices. Because CKD can develop without symptoms, it is recommended that you ask your doctor to test you for CKD at least once a year.



#### High Blood Pressure (hypertension)

High blood pressure is a major cause of kidney disease. It can cause damage to the tiny blood vessels in your kidneys. Over time, this can lead to kidney disease or kidney failure. Working with your doctor to control your blood pressure, eating a kidney-friendly diet and exercising regularly can help. Because CKD can develop without symptoms, annual screening for CKD is important as early diagnosis and treatment can help prevent CKD from progressing. It is recommended that your doctor to test you for CKD at least once a year.



#### Overweight or Obesity

If you are overweight or obese, you have a greater chance of developing diabetes or high blood pressure, which are the major causes of kidney disease and kidney failure.

PRINT YOUR RESULTS

EMAIL YOUR RESULTS

### To learn more about CKD testing

Ask your doctor for the estimated Glomerular Filtration Rate (eGFR) & Albumin-creatinine ratio, urine (ACR) tests.

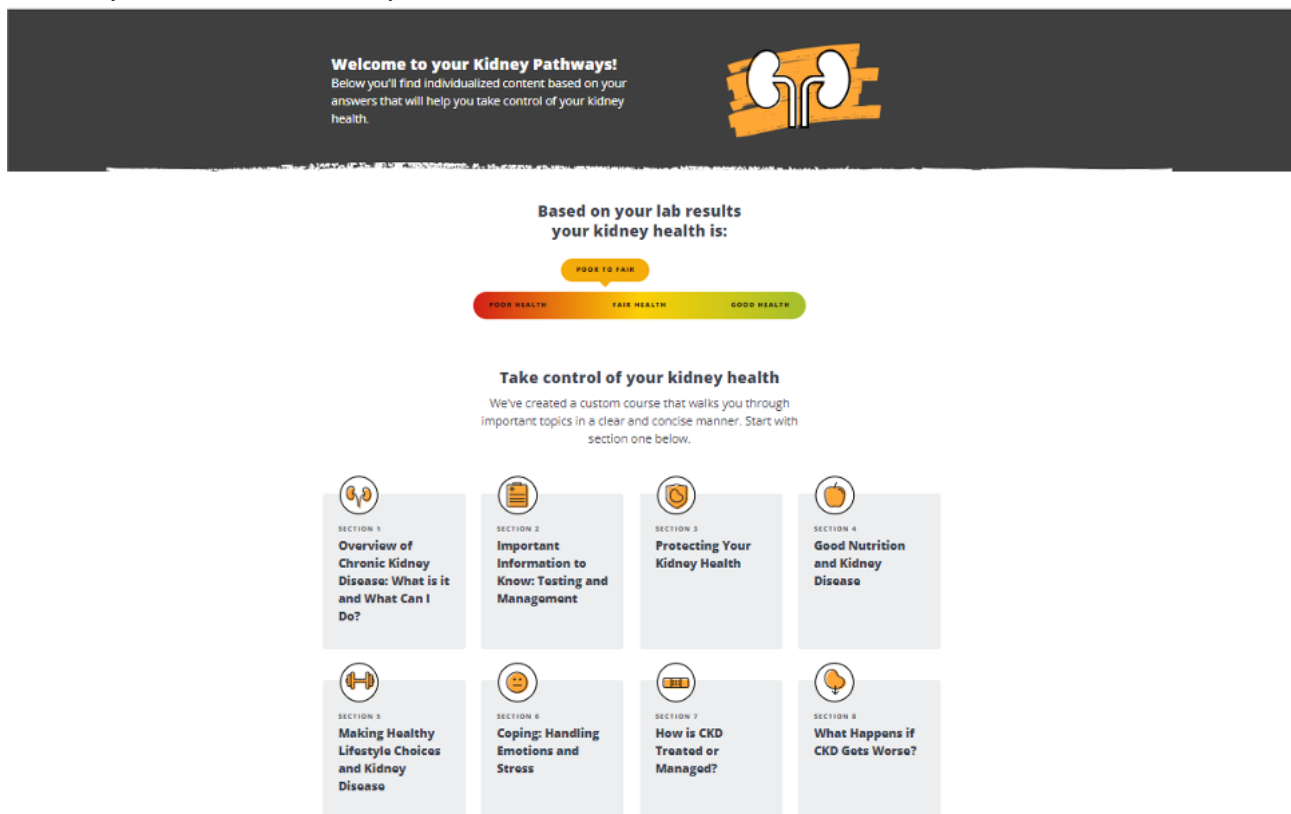
CKD is diagnosed using two, simple, inexpensive laboratory tests: **estimated Glomerular Filtration Rate** and **Albumin-creatinine ratio**, urine. Both tests are necessary to have a clear picture of your kidney health.

CKD has a silent onset, which means there are often no symptoms until the kidneys are badly damaged. Kidney tests will tell you how your kidneys are working. You should find out as early as possible if you have kidney disease to avoid problems and slow the loss of kidney function.

Blood and urine tests show how well the kidneys are doing their job. Blood and urine tests reflect the level of function of the kidneys. Urine tests can show how well the kidneys remove body wastes and whether they are leaking too much protein.

#### More about the lab tests

**Figure 3.** Patient educational materials to promote CKD self-management. This page is available to individuals who document that they are aware of their own kidney disease. CKD: chronic kidney disease.



## Participant Selection

The Kidney Score Platform was field tested in 2 phases among 20 veterans and 19 clinicians from the VA NY Harbor Healthcare System (VA-NYHHS) and the VA CT Healthcare System at West Haven (VA-CTHS). We used a purposeful sampling approach using the electronic medical record to identify potential veteran participants who were English-speaking, active primary care patients between the ages of 18 and 75 years and who were living with diabetes or hypertension, the 2 most common causes of chronic kidney disease in the United States. Although having kidney disease was not an inclusion or exclusion criterion, we excluded veterans with very advanced kidney disease, including those with an estimated glomerular filtration rate (eGFR)  $<15$  ml/min/1.73m<sup>2</sup>, those receiving dialysis treatments, and individuals who were kidney transplant recipients. Veterans who were unable to use a tablet or computer device (ie, blind, illiterate, with moderate-to-severe dementia) were also excluded. Potential veteran participants were mailed a flyer about the study and provided a phone number to opt out of the study. A research coordinator subsequently called veterans who did not opt out of the research to explain study goals and procedures, obtain consent, arrange for hard signatures of necessary forms, and schedule the think-aloud interviews.

Clinician participants included physicians and nurse practitioners who were actively engaged in primary care delivery at either Veterans Administration site. Recruitment of clinicians occurred via email by members of the research team. A research

coordinator then followed up with eligible clinicians who responded favorably to schedule the think-aloud interviews.

## Ethical Considerations

After providing online documentation of informed consent, each participant joined an online platform using a numeric study ID. Two usability facilitators (one expert consultant external to the research team and DST, a nephrologist) introduced themselves and provided a brief overview of the goals of the project. Participants were asked again to acknowledge informed consent that participation was voluntary, could cease at any time, and that the session would be audiotaped but their privacy safeguarded. The study was approved by the institutional review boards at VA-NYHHS (approval #1705) and VA-CTHS (approval #02290). The COVID-19 pandemic required protocol modifications with appropriate institutional review board amendments to allow remote participation of veterans and clinicians in online think-aloud sessions in contrast to the original investigation design, which included face-to-face study interactions. The COVID-19 pandemic also extended the research timeline, creating 2 similar phases of work instead of 1.

## Think-Aloud Testing Protocol

Each session began with a warm-up interview exploring participants' experience with online resources related to health and kidney health. Veterans were asked about experiences using web-based resources to gain information about their own health; clinicians were asked about use of online resources to communicate about chronic diseases, including but not limited to kidney disease. Thereafter, all participants were provided a

weblink to the Kidney Score Platform website and were asked to participate in a think-aloud exercise, during which they were asked to engage with the platform while verbalizing their thoughts and making their perceptions, reasonings, and decision points explicit. Think-aloud exercises are increasingly being used to conduct user testing of digital health interventions [15]. During the exercise, the usability facilitators observed participants' behaviors and probed selectively to clarify their comprehension of the tool's instructions, content, and experience. The facilitators also prompted participants to comment on the positive and negative aspects of the tool's content, language, and overall functionality.

During phase 1 (March 2020 to April 2020), audio recordings of field tests with 10 veterans and 19 clinicians were reviewed. Thematic analysis was performed, with focus on positive attributes, negative comments, and areas that required facilitator involvement. Clinicians were also asked how they might use this tool in clinical practice. Areas of improvement in the layout or design of the tool were quickly addressed by the research team and the NKF, which developed a second iteration of the Kidney Score Platform. Examples of changes included the following: reducing white space in between educational sections, rearranging location of kidney graphics, and replacing a picture of a heart with the word "love" for clarity. The Kidney Score Platform was reviewed by an additional 10 veterans during phase 2 of field testing (February 2021 to March 2021). Results from both phases were used to finalize the tool.

## Data Analysis

No formal hypothesis testing was performed due to the qualitative study methods of usability testing. Participants' demographic characteristics are described using counts and percentages. Notes from direct (although online) observations by the usability facilitators and audio recordings of the field tests were reviewed. Thematic analysis was performed (without any special software), with a focus on themes derived from the data regarding positive attributes, negative comments, and areas that required facilitator involvement.

## Results

### Participant Characteristics

Veterans were 78% (14/18) male with a mean age of 58.1 years (range 27-71 years); see Table 1. Over 60% (11/18) of veterans self-identified as African American and 6% (1/18) were Hispanic. Nearly one-third (5/18) had laboratory evidence of CKD without a formal diagnosis of CKD by problem list and 50% (9/18) carried a diagnosis of diabetes. Clinicians were 29% male (5/17). Approximately 18% (3/17) were African American, 12% (2/17) were Asian, and 65% (11/17) self-identified as non-Hispanic White. Most (15/17, 88%) were physicians (vs advanced practice providers) with a mean of 17 (range 4-32) years of experience.

**Table 1.** Characteristics of study participants.

Characteristics	Veterans (N=18)	Clinicians (N=17) <sup>a</sup>
Male gender, n (%)	14 (78)	5 (29)
Age, mean (range)	58 (27-71)	N/A <sup>b</sup>
<b>Race/ethnicity, n (%)</b>		
African American	11 (61)	3 (18)
Hispanic	1 (6)	0
Asian	0 (0)	2 (12)
Non-Hispanic White	6 (33)	11 (65)
<b>Comorbid conditions</b>		
Chronic kidney disease (eGFR <sup>c</sup> 30-60 ml/min/1.73m <sup>2</sup> ) but no diagnosis	5 (28)	N/A
Diagnosis of diabetes	9 (50)	N/A
Diagnosis of hypertension	9 (50)	N/A
<b>Practitioner type</b>		
Nurse practitioner	N/A	2 (12)
Physician	N/A	15 (88)
Time since completing training (years), mean (range)	N/A	17 (4-32)

<sup>a</sup>Data are missing for 2 clinicians who did not respond to demographic questions.

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

<sup>c</sup>eGFR: estimated glomerular filtration rate.

## Usability Testing Results

### Overview

All veterans and clinicians successfully navigated to the Kidney Score Platform website after being provided a link to the home page. Thereafter, they easily followed the self-assessment questionnaire portion of the Kidney Score Platform, aided by standard button labels and a simple progress indicator (eg, “3 of 8”) that provided a clear signal about survey length, which was also considered reasonable. The visual design of the personalized results page was appealing to all participants, who also appreciated the inclusion of infographics. Clinicians and veterans both recommended to reduce the overall text copy and include links to additional educational resources on the personalized educational results page.

Three major themes related to content and communication about risk for CKD emerged from the think-aloud exercises: (1) tension between lay and medical terminology when discussing kidney disease and diagnostic tests, (2) importance of linking general information to concrete self-management actions, and (3) usefulness of the tool as an adjunct to the office visit to prepare for patient-clinician communication. Importantly, these themes were consistent among interviews involving both veterans and clinicians.

### **Tension Between Lay and Medical Terminology When Discussing Kidney Disease Risk**

Clinicians affirmed that phrasing of questions such as “Have you been told” or “Do you take medications for [disease name]” were likely to yield accurate responses during self-assessment rather compared to asking patients whether they have a particular disease. Providers were concerned that patients may not associate (or identify) with heart disease or high blood pressure even though they were actively taking medications to manage these conditions, especially if these chronic conditions were being well managed. However, a question asking about “Have you been told” directly relates to prior conversations about chronic diseases, regardless of management strategy or success.

Aligned with clinician feedback, veterans generally understood the phrasing “Have you been told you have [health condition]” and related questions appeared to garner accurate responses from patients regarding their health risks. Veterans correctly answered yes to being told they had high blood pressure or heart failure even if they were taking medications to manage the underlying condition. However, some individuals struggled to assess whether being at risk for diabetes meant they had prediabetes, in part because they had not heard of the term “prediabetes” before.

*So my doctor never really told me I have prediabetes. He told me my A1C. [Veteran 2]*

Nearly all clinicians were concerned about veterans not recognizing the medical terms “heart failure” or “chronic kidney disease.” Clinicians stated that during their communications with patients, they were more likely to describe the underlying disease rather than rely on disease names, for example, noting to patients that “your kidneys are not functioning properly” rather than using the term “chronic kidney disease.” To be

consistent with the lay terminology used during clinic conversations, one primary care clinician suggested the following:

*I wonder if a question like, have you been told that you have chronic kidney disease, whether it should also say something like, have you been told that you have any problems with your kidneys or if your kidneys don't work completely normally or something that might not be a phrase in medical jargon. [Clinician F]*

### **Tension Between Lay and Medical Terminology When Discussing Kidney Health Tests**

Two questions embedded in the self-assessment tool asked about eGFR and urine albumin-creatinine ratio (uACR). Clinicians doubted that any of their patients would know their eGFR or uACR.

*I'm almost sure no one would know their GFR ... I think only a very special person would know [their uACR] even more so than GFR. [Clinician B]*

Most clinicians noted that they rarely used these terms in clinical practice when discussing kidney disease with their patients. Instead, they relied on plain language descriptions (“urine test” for uACR) and simplified conceptual explanations (“kidneys functioning at 60%” to describe eGFR). Although several providers acknowledged that this was an imprecise translation, they thought that it was important to make the information more accessible to patients to promote CKD awareness.

*The question that I get most often when I talk about CKD is, “What percent of my kidneys are still functioning?” Because GFR is a hard number to remember ... I'll tell people, “Listen, I think we lost 50% of your kidney function.” [Clinician G]*

Clinicians had a similar approach when discussing the presence of albuminuria.

*When I talk to patients about having protein in their urine, I don't reference the number all that often. Though more so than with eGFR, I'll pull up the trend ... But even then, I'll rarely refer to it by name. [Clinician F]*

As clinicians predicted, hardly any veterans were familiar with eGFR or uACR.

*This is like a foreign language to me. I've never heard those words before. [Veteran 3]*

For most patients, the terms themselves were new; all patients answered “I don't know” to the questions asking about levels of eGFR and uACR (Figure 2). Even among the few veterans whose doctors has spoken to them about albumin or protein in their urine, none knew the laboratory test by its clinical name, and none knew their results beyond whether they were in a normal range.

*I may not know what my numbers are, but I do know what the tests are, and I do know that I've had them done before. [Veteran 10]*

Patients used lay terminology to describe eGFR and uACR, consistent with what was described by clinicians. Most veterans said that their primary care clinician would talk more about the significance of the laboratory results as opposed to using specific terms or values.

*My doctor just said, "You don't have protein in your urine." I don't know the number or whether it was elevated. [Veteran 8]*

*I know that they have done urine tests in the past, and I know protein and sugar was in my urine. [Veteran 13]*

### **Desire for More Explicit Linkage to Self-Management Tools**

Although clinicians appreciated the personalized nature of the individualized results page about risk factors for CKD (Figure 2), they felt that there were too few actionable items that would help patients tangibly improve their health.

*I don't see much here for how to manage the risk factors... There's not much here about next steps. [Clinician F]*

*If I were a patient, I'd like links to more info about kidney-friendly eating and exercising options. I would want to have information about medications to avoid, medications that can be helpful. [Clinician C]*

Similarly, while most patients understood from the results page that they were at risk for kidney disease, some did not read carefully enough to fully realize the perceived-risk concept between diabetes, heart disease, and kidney disease and that management of the diabetes and heart disease would help mitigate risk of kidney disease. Of those patients who read the results more carefully, some felt empowered by the information, while others reacted with alarm to the risk of CKD. In particular, veterans emphasized the importance of providing actionable education to help motivate individuals to change their personal risks for kidney disease without paralyzing them with the idea that kidney failure requiring dialysis or transplant was inevitable.

*It's kind of gloom and doom; if I hit these thresholds, things may not be working properly, or may be approaching failure. So the message I get is: if you see these numbers, you're in deep trouble. [Veteran 3]*

Although most veterans planned on speaking with their clinicians about the questionnaire results, some were seeking more actionable steps that they could take on their own. They yearned for more concrete recommendations that they could adopt.

*It suggests exercise, but it's not telling me the type of exercise to do. [Veteran 1]*

*What I would be mostly interested in is what is happening, why is it happening, and what can I do to slow it down? Anything I see that's clickable that touches on those points, I would be interested in clicking on. [Veteran 11]*

### **Usefulness of the Tool as an Adjunct to the Office Visit**

When asked whether they would use the tool in clinical practice, most clinicians viewed the tool's primary value was for educating patients prior to their next appointment. With limited time during office visits, clinicians did not think that they could review the tool in its entirety with patients; however, they viewed the tool as one that patients could use in preparation for an office visit that could be dedicated to a discussion about kidney disease and cardiovascular risk. Clinicians felt that patients would benefit from having time to go through the self-assessment and results on their own time, particularly if there were more actionable next steps identified for them.

*If they can sort of generate the results and then bring them in, I think that would be helpful to have a discussion about where they are with their CKD and how we can help sort of reduce their risk of progression. [Clinician A]*

*From my perspective, I wouldn't use our visit time to go through this. Though maybe patients might find value in going through ahead of time. [Clinician H]*

Overall, most veterans felt the tool was useful since it made them aware of kidney disease. All participants said that they planned to email their CKD risk results to themselves to prepare for a discussion about kidney health with their clinician at their next visit.

*Actually, I like the site better than I thought I would. ... those questions were very precise and specific questions. As long as everybody's being honest when they're answering, I think the information that they're going to receive is going to be very useful. [Veteran 10]*

*There's some good information here. It sounds like it at least presents [information] to you a way to have this discussion with your doctor, and then see if they can test your blood or test your urine, and do the necessary tests to see where you are. [Veteran 5]*

Although no patients knew their kidney-health lab values, most were intrigued enough that they planned to bring it up with their primary care clinician at their next visit.

*It made me more interested in getting a test to see where I'm at. [Veteran 4]*

For many, being asked about their kidney health values and not knowing the answer motivated them to speak to their clinician about CKD.

*Well, let me put this way: I'm now well aware now of the significance of the kidneys and about what the issues are here. And I would definitely consider... When I go to the doctor, I would say to him, "Now, listen. You did the blood tests. But how are my kidneys doing? What are the numbers?" [Veteran 6]*

*Part of me is kind of mad there, because this a blind spot that me and my doctor, who I feel pretty comfortable with; we have not talked about CKD. I don't know if he didn't want to scare me, or maybe because he's concentrated on the prediabetes. I feel*



*like I'm going to harass him about CKD now. [Veteran 7]*

*This is something new, so immediately I was like, just another thing to be concerned about. But then I felt kind of empowered, and like I really do want to get ahead of this thing. I feel like I do want to have a conversation with my primary care physician. CKD makes me feel better than end stage renal failure, so that makes me feel empowered because I'm at risk for chronic kidney disease, which is not end stage renal failure. [Veteran 7]*

## Discussion

### Principal Findings and Comparison With Prior Work

The Kidney Score Platform is an online educational tool that was developed to promote communication and discussion about kidney disease among patients and their primary care clinicians. Usability findings demonstrate that this goal was met, with the majority of patients finding the digital tool to be helpful and easy to navigate. Content areas that would benefit from refinement were also clearly identified.

The think-aloud exercises identified 2 key themes pertinent to the development of all educational materials related to kidney disease. First, there is a tension between using medical terminology that is used regularly by clinicians as well as reinforced in laboratory reports and electronic records versus lay terminology to educate patients and promote CKD awareness. This has been an area of debate in the nephrology field for quite some time [16]. Prevalence of CKD awareness among patients differs when asked with different terms (eg, “kidney problem,” “chronic kidney disease,” “weak or failing kidneys”) [17-19]. Weak or failing kidneys is currently used in the most widely cited metric for CKD awareness in the United States, implying that low awareness is at least in part the result of semantics. Use of lay terms to describe CKD (“kidney problem”), eGFR (“percentage of kidneys filtering well”), and uACR (“protein in the urine”) in educational tools may reinforce communication about kidney disease that occurs during clinical encounters.

The use of low grade-level vocabulary is an important component of adult education and written education materials [20]. However, lay terminology is nonspecific and poses challenges for clinicians to describe the pathophysiology underlying kidney disease and to share with patients how behavioral interventions or medications will help slow kidney disease decline [21]. Complicating matters, eGFR and uACR laboratory results that are often shared with patients through patient portals exclusively use medical terminology. Paradoxically, educational tools that do not use similar medical vocabulary may thus complicate discussions about kidney disease due to disparate terminology.

Prior investigations that assess CKD awareness using lay terms such as “kidney problem” have shown incremental improvement but overall residual low awareness [17]. In an effort to increase this awareness and in response to usability testing, NKF decided to keep the eGFR and uACR terms in the Kidney Score Platform

self-assessment tool, recognizing that many patients would not know the names of these diagnostic tests. The terms were introduced with simple descriptive phrases—for example “albumin-creatinine ratio (uACR) or a type of protein in the urine”—to share the medical terminology that is used by clinical laboratories and national awareness campaigns while promoting discussions between patients and their clinicians about the usefulness and importance of these diagnostic tests to identify risks of CKD. Testing these changes in a real-world environment will demonstrate whether this strategy outweighs the risks of alienating individuals with medical terminology, particularly those who may be in denial about their chronic health conditions that increase the risk of developing CKD.

Connecting medical and lay terms may also require examples or approaches to enhance understanding. Conceptually, describing medical terminology like eGFR in ml/min/1.73m<sup>2</sup> as a percentage of kidney function, as suggested by this investigation, may help patients understand the relationship between medical and lay terms. Using the test percent performance concept for percent of kidney function with 60 or higher being normal may also make conceptual sense to patients, as 60% is approximately the customary level for scholarly examination failure in the United States. In addition, some patients may better conceptualize eGFR and uACR with the images that were explored in this investigation, with heat map colors (from low risk) of green, yellow, orange, and (high risk) red. Study findings are hypothesis generating for future investigations regarding methods to connect medical and lay terminology as well as integration of images to illustrate the relationships.

The second key theme identified the importance of linking education about kidney disease with action-oriented recommendations that can decrease the risk of kidney disease or CKD progression. Clinicians and public health officials may consider awareness of kidney disease an important outcome on its own. However, patients may consider awareness of chronic conditions only as a means to an end—a transitory step that may not lead to improved health unless the education is coupled with actionable risk-reduction tools and motivational interviewing [22]. Empowering patients with concrete examples of self-care activities that they can discuss with their clinician may help overcome the negative association between CKD awareness, poor control of CKD risk factors, and adverse health outcomes [23,24]. In direct response to this usability study, the NKF revised the “Results” page of the Kidney Score Platform. This page now has streamlined verbiage explaining the recommended tests to detect CKD, language that encourages users to read about their risk factors for kidney disease even if their underlying conditions are under control, more visible links for patients to read about actions they can take to reduce their risk of kidney disease, and additional links related to kidney-friendly diets and healthy lifestyle choices.

Both content and sequence are important elements of CKD educational media design. Workflow for time-constrained primary care clinicians is a major barrier to clear and effective bidirectional communication about kidney health, kidney disease risk, and the interplay with an array of cardiometabolic risk

conditions. We designed the study to occur before the veteran-practitioner encounters to address this barrier that in turn contributed to the clinicians recommending that the Kidney Score Platform education be delivered before the visits. Additionally, the platform could also be used after a clinician visit, allowing a modified education tailored to the interventions emphasized in the encounter or the after-visit summary for reinforcement.

### Limitations

The results of this study should be taken in context of its limitations. Although our sample size was larger than the recommended range of 5 to 7 participants for usability testing [25], our sample size was relatively small, which limits the inferences we can draw from our findings. Participants were recruited from 2 Veterans Administration clinics, which may

not be representative of primary care providers or the general adult population at risk for kidney disease nor those who deliver or receive care in nonintegrated health care delivery systems. Furthermore, all veterans were native English speakers; results cannot be generalized to populations with limited English proficiency at risk for kidney disease, for whom more targeted tools and interventions may be needed to bridge the communication divide about kidney disease.

### Conclusions

Information derived from this usability study enabled the NKF to strengthen the Kidney Score Platform tool to promote its usefulness as an empowering adjunct to care and provided some key themes that will be applicable to the development of future educational materials.

### Acknowledgments

Lesley Hunter, Gary Bronz, and Anne Cordato provided operational and administrative support. Centralis provided facilitation for the usability studies.

### Conflicts of Interest

None declared.

### References

1. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 2004 Sep 23;351(13):1296-1305. [doi: [10.1056/nejmoa041031](https://doi.org/10.1056/nejmoa041031)]
2. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. *JAMA* 2002 Nov 20;288(19):2469-2475. [doi: [10.1001/jama.288.19.2469](https://doi.org/10.1001/jama.288.19.2469)] [Medline: [12435261](https://pubmed.ncbi.nlm.nih.gov/12435261/)]
3. Chu CD, McCulloch CE, Banerjee T, Pavkov ME, Burrows NR, Gillespie BW, Centers for Disease ControlPrevention Chronic Kidney Disease Surveillance Team. CKD awareness among US Adults by future risk of kidney failure. *Am J Kidney Dis* 2020 Aug;76(2):174-183 [FREE Full text] [doi: [10.1053/j.ajkd.2020.01.007](https://doi.org/10.1053/j.ajkd.2020.01.007)] [Medline: [32305206](https://pubmed.ncbi.nlm.nih.gov/32305206/)]
4. Tuot DS, Plantinga LC, Hsu C, Jordan R, Burrows NR, Hedgeman E, et al. Chronic kidney disease awareness among individuals with clinical markers of kidney dysfunction. *CJASN* 2011 Jul 22;6(8):1838-1844. [doi: [10.2215/cjn.00730111](https://doi.org/10.2215/cjn.00730111)]
5. Chao C, Lee Y, Yang K, Peng J, Li C, Chen S, Cogent Study Group. Impact of self-report and eGFR-based chronic kidney disease on the risk of chronic kidney disease-related complications and geriatric syndromes in community-dwelling older adults. *Kidney Blood Press Res* 2018 Dec 14;43(6):1908-1918 [FREE Full text] [doi: [10.1159/000496002](https://doi.org/10.1159/000496002)] [Medline: [30566955](https://pubmed.ncbi.nlm.nih.gov/30566955/)]
6. Greer R, Crews DC, Boulware LE. Challenges perceived by primary care providers to educating patients about chronic kidney disease. *J Ren Care* 2012 Dec;38(4):174-181 [FREE Full text] [doi: [10.1111/j.1755-6686.2012.00323.x](https://doi.org/10.1111/j.1755-6686.2012.00323.x)] [Medline: [23176576](https://pubmed.ncbi.nlm.nih.gov/23176576/)]
7. Shin JI, Chang AR, Grams ME, Coresh J, Ballew SH, Surapaneni A, et al. Albuminuria testing in hypertension and diabetes: an individual-participant data meta-analysis in a global consortium. *Hypertension* 2021 Sep;78(4):1042-1052 [FREE Full text] [doi: [10.1161/HYPERTENSIONAHA.121.17323](https://doi.org/10.1161/HYPERTENSIONAHA.121.17323)] [Medline: [34365812](https://pubmed.ncbi.nlm.nih.gov/34365812/)]
8. Lee J, Chu C, Guzman D, Fontil V, Velasquez A, Powe N, et al. Albuminuria testing by race and ethnicity among patients with hypertension with and without diabetes. *Am J Nephrol* 2019;50(1):48-54 [FREE Full text] [doi: [10.1159/000500706](https://doi.org/10.1159/000500706)] [Medline: [31167180](https://pubmed.ncbi.nlm.nih.gov/31167180/)]
9. Greer RC, Cooper LA, Crews DC, Powe NR, Boulware LE. Quality of patient-physician discussions about CKD in primary care: a cross-sectional study. *Am J Kidney Dis* 2011 Apr;57(4):583-591 [FREE Full text] [doi: [10.1053/j.ajkd.2010.08.027](https://doi.org/10.1053/j.ajkd.2010.08.027)] [Medline: [21131116](https://pubmed.ncbi.nlm.nih.gov/21131116/)]
10. Sperati CJ, Soman S, Agrawal V, Liu Y, Abdel-Kader K, Diamantidis CJ, et al. Primary care physicians' perceptions of barriers and facilitators to management of chronic kidney disease: a mixed methods study. *PLoS ONE* 2019 Aug 22;14(8):e0221325. [doi: [10.1371/journal.pone.0221325](https://doi.org/10.1371/journal.pone.0221325)]
11. van Dulmen S, Roodbeen R, Schulze L, Prantl K, Rookmaaker M, van Jaarsveld B, et al. Practices and perspectives of patients and healthcare professionals on shared decision-making in nephrology. *BMC Nephrol* 2022 Jul 21;23(1):258 [FREE Full text] [doi: [10.1186/s12882-022-02887-4](https://doi.org/10.1186/s12882-022-02887-4)] [Medline: [35864466](https://pubmed.ncbi.nlm.nih.gov/35864466/)]

12. Boulware LE, Carson KA, Troll MU, Powe NR, Cooper LA. Perceived susceptibility to chronic kidney disease among high-risk patients seen in primary care practices. *J Gen Intern Med* 2009 Oct;24(10):1123-1129 [FREE Full text] [doi: [10.1007/s11606-009-1086-6](https://doi.org/10.1007/s11606-009-1086-6)] [Medline: [19711135](https://pubmed.ncbi.nlm.nih.gov/19711135/)]
13. Tuot DS, Crowley ST, Katz LA, Leung J, Alcantara-Cadillo DK, Ruser C, et al. The Kidney Score pPlatform for patient and clinician awareness, communication, and management of kidney disease: protocol for a mixed methods stud. *JMIR Res Protoc* 2020 Oct 19;9(10):e22024 [FREE Full text] [doi: [10.2196/22024](https://doi.org/10.2196/22024)] [Medline: [33074162](https://pubmed.ncbi.nlm.nih.gov/33074162/)]
14. Vassalotti JA, Boucree SC. Integrating CKD into US primary care: bridging the knowledge and implementation gaps. *Kidney Int Rep* 2022 Mar;7(3):389-396 [FREE Full text] [doi: [10.1016/j.ekir.2022.01.1066](https://doi.org/10.1016/j.ekir.2022.01.1066)] [Medline: [35257053](https://pubmed.ncbi.nlm.nih.gov/35257053/)]
15. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. *Int J Med Inform* 2019 Jun;126:95-104. [doi: [10.1016/j.ijmedinf.2019.03.018](https://doi.org/10.1016/j.ijmedinf.2019.03.018)] [Medline: [31029270](https://pubmed.ncbi.nlm.nih.gov/31029270/)]
16. Levey A, Eckardt K, Dorman N, Christiansen S, Cheung M, Jadoul M, et al. Nomenclature for kidney function and disease: report of a Kidney Disease: Improving Global Outcomes (KDIGO) Consensus Conference. *Kidney Dis (Basel)* 2020 Sep;6(5):309-317 [FREE Full text] [doi: [10.1159/000509359](https://doi.org/10.1159/000509359)] [Medline: [33490111](https://pubmed.ncbi.nlm.nih.gov/33490111/)]
17. Tuot DS, Zhu Y, Velasquez A, Espinoza J, Mendez CD, Banerjee T, et al. Variation in patients' awareness of CKD according to how they are asked. *CJASN* 2016 Jun 23;11(9):1566-1573. [doi: [10.2215/cjn.00490116](https://doi.org/10.2215/cjn.00490116)]
18. Tuot D, Wong KK, Velasquez A, Crews DC, Zonderman AB, Evans MK, et al. CKD awareness in the general population: performance of CKD-specific questions. *Kidney Med* 2019;1(2):43-50 [FREE Full text] [doi: [10.1016/j.xkme.2019.01.005](https://doi.org/10.1016/j.xkme.2019.01.005)] [Medline: [32734184](https://pubmed.ncbi.nlm.nih.gov/32734184/)]
19. Chu CD, Chen MH, McCulloch CE, Powe NR, Estrella MM, Shlipak MG, et al. Patient awareness of CKD: a systematic review and meta-analysis of patient-oriented questions and study setting. *Kidney Med* 2021;3(4):576-585.e1 [FREE Full text] [doi: [10.1016/j.xkme.2021.03.014](https://doi.org/10.1016/j.xkme.2021.03.014)] [Medline: [34401725](https://pubmed.ncbi.nlm.nih.gov/34401725/)]
20. Tuot DS, Davis E, Velasquez A, Banerjee T, Powe NR. Assessment of printed patient-educational materials for chronic kidney disease. *Am J Nephrol* 2013 Aug 21;38(3):184-194 [FREE Full text] [doi: [10.1159/000354314](https://doi.org/10.1159/000354314)] [Medline: [23970127](https://pubmed.ncbi.nlm.nih.gov/23970127/)]
21. Palazzuoli A, Tecson KM, Ronco C, McCullough PA. Nomenclature for kidney function from KDIGO: shortcomings of terminology oversimplification. *Cardiorenal Med* 2021;11(3):119-122 [FREE Full text] [doi: [10.1159/000516615](https://doi.org/10.1159/000516615)] [Medline: [34091445](https://pubmed.ncbi.nlm.nih.gov/34091445/)]
22. Loskutova N, Tsai AG, Callen E, Ajayi K, Carroll JK, Harrington M, et al. Differences in perspectives regarding diabetes management between health care providers and patients. *Transl Behav Med* 2018 May 23;8(3):328-340. [doi: [10.1093/tbm/iby024](https://doi.org/10.1093/tbm/iby024)] [Medline: [29800415](https://pubmed.ncbi.nlm.nih.gov/29800415/)]
23. Tuot DS, Plantinga LC, Judd SE, Muntner P, Hsu C, Warnock DG, REGARDS Investigators. Healthy behaviors, risk factor control and awareness of chronic kidney disease. *Am J Nephrol* 2013;37(2):135-143 [FREE Full text] [doi: [10.1159/000346712](https://doi.org/10.1159/000346712)] [Medline: [23392070](https://pubmed.ncbi.nlm.nih.gov/23392070/)]
24. Tummalapalli S, Vittinghoff E, Crews D, Cushman M, Gutiérrez OM, Judd S, et al. Chronic kidney disease awareness and longitudinal health outcomes: results from the reasons for geographic and racial differences in stroke study. *Am J Nephrol* 2020;51(6):463-472 [FREE Full text] [doi: [10.1159/000507774](https://doi.org/10.1159/000507774)] [Medline: [32349001](https://pubmed.ncbi.nlm.nih.gov/32349001/)]
25. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. *Behav Res Methods Instrum Comput* 2003 Aug;35(3):379-383. [doi: [10.3758/bf03195514](https://doi.org/10.3758/bf03195514)] [Medline: [14587545](https://pubmed.ncbi.nlm.nih.gov/14587545/)]

## Abbreviations

**CKD:** chronic kidney disease eGFR: estimated glomerular filtration rate

**NKF:** National Kidney Foundation

**uACR:** urine albumin-creatinine ratio

**VA-NYHHS:** VA NY Harbor Healthcare System

**VA-CTHS:** VA CT Healthcare System at West Haven

*Edited by A Mavragani; submitted 31.05.22; peer-reviewed by S Song, D Nash; comments to author 18.07.22; revised version received 02.09.22; accepted 02.09.22; published 28.09.22.*

*Please cite as:*

Tuot DS, Crowley ST, Katz LA, Leung J, Alcantara-Cadillo DK, Ruser C, Talbot-Montgomery E, Vassalotti JA

*Usability Testing of the Kidney Score Platform to Enhance Communication About Kidney Disease in Primary Care Settings: Qualitative Think-Aloud Study*

*JMIR Form Res* 2022;6(9):e40001

URL: <https://formative.jmir.org/2022/9/e40001>

doi: [10.2196/40001](https://doi.org/10.2196/40001)

PMID: [36170008](https://pubmed.ncbi.nlm.nih.gov/36170008/)

©Delphine S Tuot, Susan T Crowley, Lois A Katz, Joseph Leung, Dely K Alcantara-Cadillo, Christopher Ruser, Elizabeth Talbot-Montgomery, Joseph A Vassalotti. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 28.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Comparison Between QT and Corrected QT Interval Assessment by an Apple Watch With the AccurBeat Platform and by a 12-Lead Electrocardiogram With Manual Annotation: Prospective Observational Study

Sara Chokshi<sup>1</sup>, DrPH; Gulzhan Tologonova<sup>2</sup>, MD; Rose Calixte<sup>3</sup>, PhD; Vandana Yadav<sup>1</sup>, MSc; Naveed Razvi<sup>4</sup>, MD; Jason Lazar<sup>2</sup>, MPH, MD; Stan Kachnowski<sup>1,5,6</sup>, MPA, PhD

<sup>1</sup>Healthcare Innovation and Technology Lab, New York, NY, United States

<sup>2</sup>Division of Cardiovascular Medicine, State University of New York Downstate Medical Center, New York, NY, United States

<sup>3</sup>Department of Epidemiology and Biostatistics, State University of New York Downstate Health Sciences University, New York, NY, United States

<sup>4</sup>Department of Cardiology, Ipswich Hospital, Ipswich, United Kingdom

<sup>5</sup>Columbia Business School, Columbia University, New York, NY, United States

<sup>6</sup>Indian Institute of Technology Delhi, Delhi, India

**Corresponding Author:**

Sara Chokshi, DrPH

Healthcare Innovation and Technology Lab

3960 Broadway

New York, NY, 10032

United States

Phone: 1 212 543 0100

Email: [schokshi@hitlab.org](mailto:schokshi@hitlab.org)

## Abstract

**Background:** Abnormal prolongation or shortening of the QT interval is associated with increased risk for ventricular arrhythmias and sudden cardiac death. For continuous monitoring, widespread use, and prevention of cardiac events, advanced wearable technologies are emerging as promising surrogates for conventional 12-lead electrocardiogram (ECG) QT interval assessment. Previous studies have shown a good agreement between QT and corrected QT (QTc) intervals measured on a smartwatch ECG and a 12-lead ECG, but the clinical accuracy of computerized algorithms for QT and QTc interval measurement from smartwatch ECGs is unclear.

**Objective:** The prospective observational study compared the smartwatch-recorded QT and QTc assessed using AccurKardia's AccurBeat platform with the conventional 12-lead ECG annotated manually by a cardiologist.

**Methods:** ECGs were collected from healthy participants (without any known cardiovascular disease) aged >22 years. Two consecutive 30-second ECG readings followed by (within 15 minutes) a 10-second standard 12-lead ECG were recorded for each participant. Characteristics of the participants were compared by sex using a 2-sample *t* test and Wilcoxon rank sum test. Statistical comparisons of heart rate (HR), QT interval, and QTc interval between the platform and the 12-lead ECG, ECG lead I, and ECG lead II were done using the Wilcoxon sign rank test. Linear regression was used to predict QTc and QT intervals from the ECG based on the platform's QTc/QT intervals with adjustment for age, sex, and difference in HR measurement. The Bland-Altman method was used to check agreement between various QT and QTc interval measurements.

**Results:** A total of 50 participants (32 female, mean age 46 years, SD 1 year) were included in the study. The result of the regression model using the platform measurements to predict the 12-lead ECG measurements indicated that, in univariate analysis, QT/QTc intervals from the platform significantly predicted QT/QTc intervals from the 12-lead ECG, ECG lead I, and ECG lead II, and this remained significant after adjustment for sex, age, and change in HR. The Bland-Altman plot results found that 96% of the average QTc interval measurements between the platform and QTc intervals from the 12-lead ECG were within the 95% confidence limit of the average difference between the two measurements, with a mean difference of -10.5 (95% limits of agreement -71.43, 50.43). A total of 94% of the average QT interval measurements between the platform and the 12-lead ECG were within the 95% CI of the average difference between the two measurements, with a mean difference of -6.3 (95% limits of agreement -54.54, 41.94).

**Conclusions:** QT and QTc intervals obtained by a smartwatch coupled with the platform's assessment were comparable to those from a 12-lead ECG. Accordingly, with further refinements, remote monitoring using this technology holds promise for the identification of QT interval prolongation.

(*JMIR Form Res* 2022;6(9):e41241) doi:[10.2196/41241](https://doi.org/10.2196/41241)

## KEYWORDS

artificial intelligence ECG; AI ECG; AI wearables; big data; cardiovascular medicine; digital health; machine learning

## Introduction

Cardiovascular disease is highly prevalent and poses significant morbidity and mortality, accounting for approximately 1 in every 4 deaths in the United States alone [1,2]. Remote monitoring of heart health to detect early signs of deterioration and thus opportunities for intervention prior to situations requiring inpatient care is an area that would benefit from a cost-saving outcome-improving innovation.

The 12-lead electrocardiogram (ECG) has long been a standard component of evaluation for patients suspected of or confirmed to have cardiovascular disease. However, use of the 12-lead ECG is restricted to medical facilities, as qualified physicians are needed to interpret the results [3-5]. In low- and middle-income countries, where over 75% of deaths are related to cardiovascular disease and where there is limited access to ECG equipment and cardiologists, many patients with heart disease need regular ECG checks and reviews, both to check for disease progression and for surveillance of drug therapies [6]. This demand can place an insurmountable burden on the available pool of specialists, including in areas with a high concentration of populations susceptible to heart disease (eg, South Asian people), particularly in the COVID-19 pandemic era [7].

The QT interval is an important parameter derived from the 12-lead ECG that represents the time for ventricular depolarization. A key data point of interest in the ECG is the QT interval [8-10]. While most commonly corrected for heart rate (HR), abnormal prolongation or shortening of the QT interval, whether congenital [11,12] or acquired, is associated with increased risk for ventricular arrhythmias and even sudden cardiac death [13].

The QT interval is also an important parameter to follow in patients treated with cardiac medications. Advanced wearable technologies provide new opportunities for the diagnosis and management of cardiovascular diseases and their risk factors in the convenience of patients' homes and other nonclinical settings [14]. The latest generation of smartwatches and smartphones are increasingly popular tools for health monitoring and care delivery, capable of collecting key vital signs such as HR, blood pressure, and even ECG data [15]. The Apple Watch, which recently received approval for atrial fibrillation detection from the Food and Drug Administration, can perform an ECG using a single peripheral lead (lead I)—obtained through a circuit between the detector on the back of the watch and the digital crown. While the Apple Watch single-lead ECG can detect atrial fibrillation, the feasibility and diagnostic accuracy for QT interval measurement is less established [10,16]. While

previous studies have already shown a good agreement between QT and corrected QT (QTc) intervals measured on a smartwatch ECG and a 12-lead ECG [17], the use of computerized algorithms for QT and QTc interval measurement from smartwatch ECGs lacks a similar level of evidence [18]. This is a key step in fulfilling the promise of using wearable technologies to facilitate the diagnosis and management of cardiovascular health [19].

To build upon the promise of this new technology, a cardiology-focused digital health company (AccurKardia) has developed a device agnostic platform (AccurBeat) for the analysis of Apple Watch (version 4 or higher)—generated ECGs that leverages an engine built on computational and artificial intelligence (AI) techniques to perform automated analysis of ECGs and support the early detection and diagnosis of arrhythmias. The objective of this study is to compare smartwatch-recorded QT and QTc intervals assessed using the platform's algorithm with the conventional gold standard procedure that uses a 12-lead ECG annotated manually by an expert cardiologist.

## Methods

### Study Design

This study is a single-site observational study to compare QT and QTc intervals assessed using smartwatch-generated data coupled with the platform's algorithm and QT and QTc intervals measured using a 12-lead ECG with manual annotation in healthy individuals. The study was performed in the Noninvasive Cardiology Unit at the State University of New York (SUNY) Downstate Medical Center in Brooklyn, New York, a large urban medical center.

### Ethics Approval

The study was deemed as human participants research and was approved by the Biomedical Research Alliance of New York Institutional Review Board and the Institutional Review Board at SUNY Downstate Medical Center (IRB 21-02-474), and all participants provided informed consent that was delivered and documented by the study coordinator prior to data collection. All participant data was collected anonymously. Aside from the Apple Watch-generated readings, data were recorded and stored securely on SUNY Downstate Medical Center servers in password-protected spreadsheets. Apple Watch-generated data, once captured, was automatically sent through an API call to a cloud-based analytics engine ("AccurAI"), which annotated the ECG and provide a computerized interpretation of the rhythm that was identified. The output of these analytics was accessible through a secure web-based clinician portal for the AccurBeat device.

## Study Population

Healthy adult participants without known or suspected heart disease were recruited from outpatient primary care and cardiology clinics between January 6 and 19, 2022. This was a convenience sample from the SUNY Downstate cardiology clinic and internal medicine practice. Patients were screened by the study coordinator daily prior to their visit and, if qualified, were informed of the study. In total, 54 patients were screened, and 50 patients provided informed consent. The inclusion criteria were selected based on both patient self-report and electronic medical records. The exclusion criteria included any recent illness within 4 weeks and taking any medication irrespective of an indication that is known to prolong the QT interval.

## Study Procedure

### Data Collection

Each participant was informed about the study procedures, and written consent was collected. Each participant was then asked to sit down while a highly experienced study coordinator placed the study-dedicated Apple Watch (version 7) on their left wrist and facilitated two consecutive 30-second ECG readings. Within 15 minutes following the Apple Watch readings, the study coordinator had participants lie flat and proceeded to place electrodes on the participants to perform a 10-second standard 12-lead ECG reading using the GE MAC 5500HD ECG machine with a paper speed of 25 mm per second. All participants were compensated (US \$75) via gift card following their study visit.

### Data Storage and Analysis

Apple Watch ECG data was automatically uploaded to Apple HealthKit, Apple's central repository for health and fitness data on the iPhone and was confirmed immediately following collection by the study coordinator via the platform's smartphone app. The 12-lead ECG readings were printed at the time of reading and labeled with the data and time of reading for identification purposes. All Apple Watch ECG data were assessed using the platform. The corresponding 12-lead ECG was assessed and manually annotated by an expert cardiologist.

### Annotation Procedures

The ECGs were recorded on paper tracings. They were digitalized and then imported in ImageJ (free online software provided by the National Institutes of Health). The calibrations were performed for 0.4 seconds, and the followings measures were made for each beat in each lead: in QQ interval and in respiratory rate interval.

All ECGs had a placing that was technically adequate for analysis. In the case of a flattened T wave, the lead was excluded for analysis from the QT interval of the 12 leads. All analyzable complexes were in lead I and lead II. Bazett's [20] formula was used to correct for HR in all determinations.

### Solution Development and Evaluation

The AccurBeat (version 1.0) platform includes a native iOS app (used to view the annotated ECG and computerized interpretation of rhythm classification), a clinician web portal (for the review and approval of reports prior to release to patients), a cloud-based application programming interface to

access the analytics engine, and the analytics engine itself that annotates the ECG and provides a computerized interpretation of rhythm classification. The analytics engine is based on proprietary methods that leverage a combination of signal processing, image analysis, and AI-based techniques to annotate ECGs and diagnose arrhythmias. The data is normalized, and features are extracted using various signal processing techniques. Once this initial processing is complete, a hybrid architecture combining image analysis with evolutionary computing-based AI is invoked for beat classification, complex feature extraction, and rhythm detection. Following this, an inference engine with established clinical guidelines is used to obtain a diagnosis. Since this study only focused on HR, QT interval, and QTc interval measurements, the output of the inference engine was not applicable to the results of this study. The algorithm was previously tested according to the AAMI ANSI EC57:2012 standard with both publicly available and proprietary databases.

### Statistical Analysis

Quantitative variables (age, HR, QT interval, and QTc interval) are summarized as means (SD) or median (IQR), and categorical variables are reported as frequencies (percentages). We compared characteristics of the participants by sex using a 2-sample *t* test and Wilcoxon rank sum test. Statistical comparisons of HR, QT interval, and QTc interval between the platform and 12-lead ECG, ECG lead I, and ECG lead II were done using the Wilcoxon sign rank test. We used linear regression to predict the QTc and QT intervals from the ECG based on the platform's QTc/QT intervals, with adjustment for age, sex, and the difference in HR measurement. We checked for multicollinearity using variance inflation factor (VIF).

Agreement between QT and QTc interval measurements (taken on 12-lead ECGs and annotated manually and taken on smartwatches and assessed by the platform) was assessed using the Bland-Altman method [21,22]. The mean of the difference (bias) in QT and QTc intervals between the two methods was calculated, along with the 95% lower and upper limits of agreement (LoA). Agreement between the measures was also numerically assessed by estimating the agreement intraclass correlation coefficient, with its 95% CI [23]. Statistical significance was set at .05. All analyses were done in R 4.0.3 (R Foundation for Statistical Computing) and RStudio 1.2.5019 (RStudio, PBC).

## Results

In Table 1, we summarized the characteristics of the study participants and compared them by sex. Of all 50 participants, 32 (64%) of the study participants identified as female. They had a mean age of 46.18 (11.89) years. There was no sex difference in mean age, mean HR measurements from all devices, or mean QT interval measurements from all devices. However, QTc interval measurements from all devices were significantly higher for female patients compared to male patients.

In Table 2, we summarized HR, QT intervals, and QTc intervals between the platform and 12-lead ECG, ECG lead I, and ECG lead II. Results of the Wilcoxon sign rank test indicated that all

measurements from the AccurBeat device were significantly higher than those from the 12-lead ECG.

The correlations, with 95% CIs, between the platform measurements and 12-lead ECG measurements are featured in Table 3. The result indicated that the correlations between measurements across devices were all significantly different from 0. However, the strengths of association range from low to strong positive associations. In the sensitivity analysis, the results remained consistent.

The result of the regression model using the platform measurements to predict the 12-lead ECG measurements (Table 4) indicate that, in univariate analysis, QT/QTc intervals from the platform significantly predicted QT/QTc intervals from the 12-lead ECG, ECG lead I, and ECG lead II. The significant association between QT/QTc intervals from the platform and QT/QTc intervals from the 12-lead ECG remained significant after adjustment for sex, age, and change in HR. In the multivariable model, for each unit increase in the platform QTc

interval, the QTc interval from the 12-lead ECG was expected to increase by 0.31 (adjusted  $R^2=0.38$ ). Similarly, the QTc interval from ECG lead I and ECG lead II were expected to increase significantly by 0.30 (adjusted  $R^2=0.25$  and 0.32 for lead I and lead II, respectively). A 33-point increase in the platform QTc interval would correspond to approximately a 10-point increase in the QTc interval from the 12-lead ECG, adjusting for age, sex, and change in HR. In the multivariable model, for each unit increase in the platform QT interval, the QT interval from the 12-lead ECG was expected to increase by 0.53 (adjusted  $R^2=0.47$ ). Similarly, the QT interval from ECG lead I and ECG lead II were expected to increase significantly by 0.43 and 0.48 (adjusted  $R^2=0.39$  and 0.29), respectively. A 19-point increase in the platform QT interval (24 for lead I and 21 for lead II) would correspond to an approximately 10-point increase in the QT interval from the 12-lead ECG, adjusting for age, sex, and change in HR. The VIF for all six models was less than 2, indicating no multicollinearity.

**Table 1.** Descriptive summary of age, HR, QT interval, and QTc interval by sex.

	Female (n=32)	Male (n=18)	P value <sup>a</sup>
Age (years), mean (SD)	47.44 (10.99)	43.94 (13.38)	.35
Platform HR <sup>b</sup> (bpm), mean (SD)	78.67 (9.92)	74.50 (11.79)	.21
Platform QT interval (ms), median (IQR)	401 (371.8-424)	374 (369.1-392.4)	.10
Platform QTc <sup>c</sup> interval (ms), median (IQR)	444.5 (433.4-465.2)	423.8 (410.2-434.2)	<.001
12-lead HR (bpm), mean (SD)	75.22 (9.99)	73.00 (11.34)	.49
12-lead QT interval (ms), median (IQR)	382 (363.5-410.5)	375 (360-388)	.29
12-lead QTc interval (ms), median (IQR)	431.5 (412-441)	410.5 (40.32-415.8)	.006
Lead I HR (bpm), mean (SD)	77.42 (11.62)	72.17 (12.72)	.16
Lead I QT interval (ms), median (IQR)	366.5 (350.8-393.2)	363.5 (349.5-381)	.56
Lead I QTc interval (ms), median (IQR)	418 (406.5-435)	397 (373.5-718.8)	.02
Lead II HR (bpm), mean (SD)	76.68 (10.34)	72.06 (14.38)	.24
Lead II QT interval (ms), median (IQR)	371 (356-396.2)	362.5 (353.2-380.2)	.18
Lead II QTc interval (ms), median (IQR)	430 (410-436.5)	388 (366.5-410.2)	.003

<sup>a</sup>Statistical comparison between measurements from the platform and the 12-lead electrocardiogram were done using a 2-sample *t* test and Wilcoxon rank sum test. A *P* value <.05 was considered significant.

<sup>b</sup>HR: heart rate.

<sup>c</sup>QTc: corrected QT.



**Table 2.** Descriptive summary of HR, QT interval, and QTc interval (N=50).<sup>a</sup>

	Values			P value
	Mean (SD)	Range	Median (IQR)	
<b>Platform</b>				
HR <sup>b</sup> (bpm)	77.17 (10.70)	53.50-97.50	76.5 (68.12-84.88)	N/A <sup>c</sup>
QT interval (ms)	389.9 (33.95)	293.5-443.5	388.5 (369.9-417.9)	N/A
QTc <sup>d</sup> interval (ms)	434.4 (32.91)	305-487.5	437.5 (423.4-459.9)	N/A
<b>12-lead ECG<sup>e</sup></b>				
HR (bpm)	74.42 (10.44)	54-100	74 (66.5-80.75)	<.001
QT interval (ms)	383.6 (26.63)	342-442	378 (362.5-406)	.005
QTc interval (ms)	423.9 (23.16)	379-486	422.5 (408-438.8)	.001
<b>ECG lead I</b>				
HR (bpm)	75.49 (12.17)	47-98	75 (67-84)	.047
QT interval (ms)	368.9 (28.01)	318-429	365.5 (350.2-389.8)	<.001
QTc interval (ms)	411.1 (27.30)	359-481	414 (392-426)	<.001
<b>ECG lead II</b>				
HR (bpm)	74.98 (12.05)	47-105	75 (67-84)	.012
QT interval (ms)	371.5 (31.35)	293-468	370 (354-390.8)	<.001
QTc interval (ms)	412.4 (34.34)	336-494	417 (387-436)	<.001

<sup>a</sup>Statistical comparison between measurements from the platform and 12-lead ECG were done using Wilcoxon sign rank test.

<sup>b</sup>HR: heart rate.

<sup>c</sup>N/A: not available.

<sup>d</sup>QTc: corrected QT.

<sup>e</sup>ECG: electrocardiogram.

**Table 3.** Correlation between devices.

Platform measures with...	Intraclass correlation: consistency (95% CI)	Intraclass correlation: agreement (95% CI)	Pearson correlation (95% CI)
HR <sup>a</sup> from 12-lead ECG <sup>b</sup>	0.88 (0.79-0.93)	0.85 (0.69-0.92)	0.87 (0.79-0.93)
HR (lead I)	0.75 (0.59-0.85)	0.74 (0.58-0.84)	0.75 (0.60-0.85)
HR (lead II)	0.79 (0.65-0.87)	0.77 (0.62-0.87)	0.79 (0.66-0.88)
QTc <sup>c</sup> interval from 12-lead ECG	0.40 (0.14-0.61)	0.38 (0.13-0.59)	0.43 (0.17-0.63)
QTc interval (lead I)	0.42 (0.16-0.63)	0.33 (0.02-0.57)	0.43 (0.17-0.63)
QTc interval (lead II)	0.41 (0.15-0.62)	0.34 (0.05-0.57)	0.41 (0.14-0.62)
QT interval from 12-lead ECG	0.68 (0.49-0.80)	0.66 (0.48-0.80)	0.69 (0.52-0.82)
QT interval (lead I)	0.53 (0.30-0.70)	0.43 (0.09-0.66)	0.54 (0.31-0.71)
QT interval (lead II)	0.56 (0.34-0.73)	0.49 (0.17-0.70)	0.56 (0.34-0.73)

<sup>a</sup>HR: heart rate.

<sup>b</sup>ECG: electrocardiogram.

<sup>c</sup>QTc: corrected QT.

**Table 4.** Association of the platform's QT/QTc intervals with the 12-lead electrocardiogram's QT/QTc intervals.<sup>a</sup>

	12-lead		Lead I				Lead II					
	$\beta$ b,c (SE)	P value	$\beta$ b,d (SE)	P value	$\beta$ b,c (SE)	P value	$\beta$ b,d (SE)	P value	$\beta$ b,c (SE)	P value	$\beta$ b,c (SE)	P value
<b>QTc<sup>e,f</sup></b>	0.30 (0.09)	.002	0.31 (0.09)	.001	0.35 (0.11)	.002	0.30 (0.12)	.01	0.42 (0.14)	.002	0.30 (0.14)	.03
Age			0.30 (0.22)	.18			0.33 (0.29)	.26			0.04 (0.36)	.92
Male			-10.66 (6.03)	.08			-10.84 (7.87)	.26			-22.29 (9.41)	.92
HR <sup>g,h</sup>			-1.93 (0.52)	<.001			-0.93 (0.43)	.03			-1.62 (0.56)	.006
<b>QT<sup>f</sup></b>	0.54 (0.08)	<.001	0.53 (0.08)	<.001	0.44 (0.10)	<.001	0.41 (0.09)	<.001	0.52 (0.11)	<.001	0.48 (0.12)	<.001
Age			0.37 (0.24)	.13			0.47 (0.26)	.08			0.19 (0.33)	.58
Male			0.54 (6.02)	.93			2.03 (6.47)	.75			-8.13 (8.11)	.32
HR			-0.29 (0.53)	.59			1.09 (0.38)	.006			0.54 (0.53)	.31

<sup>a</sup>The QT/QTc intervals from the 12-lead electrocardiogram were modeled using multiple linear regression with QT/QTc intervals from the platform as the main predictor, adjusted for age, sex, and change in HR.

<sup>b</sup> $\beta$ : parameter estimates.

<sup>c</sup>simple linear regression.

<sup>d</sup>multiple linear regression.

<sup>e</sup>QTc: corrected QT.

<sup>f</sup>QTc/QT measurements were from the platform.

<sup>g</sup>HR: heart rate.

<sup>h</sup>Change in HR between the platform and electrocardiogram.

The Bland-Altman plot results found that 96% of the average QTc interval measurements between the platform and the QTc intervals from the 12-lead ECG were within the 95% confidence limit of the average difference between the two measurements (Figures 1-6), with a mean difference of -10.5 (95% LoA -71.43, 50.43). The Bland-Altman analysis detected a significant proportional bias between the AccurBeat QTc interval and the QTc interval from the 12-lead ECG ( $P=.008$ ). Over 95% of the average QTc interval measurements between the platform and QTc intervals from the 12-lead ECG (lead I) were within the 95% confidence limit of the average difference between the two measurements, with a mean difference of -23.45 (95% LoA -87.62, 40.72). The Bland-Altman analysis detected no significant proportional bias between the AccurBeat QTc intervals and the QTc intervals from the ECG lead I ( $P=.14$ ). Over 93% of the average QTc interval measurements between the platform and the 12-lead ECG (lead II) were within the 95% confidence limit of the average difference between the two measurements, with a mean difference of -22.2 (95% LoA -94.15, 49.82). The Bland-Altman analysis detected no significant proportional bias between the AccurBeat QTc intervals and the QTc intervals from the ECG lead II ( $P=.81$ ).

A total of 94% of the average QT interval measurements between the platform and QT intervals from the 12-lead ECG were within the 95% CI for the average difference between the two measurements, with a mean difference of -6.3 (95% LoA -54.54, 41.94). The Bland-Altman analysis detected a significant proportional bias between the AccurBeat QT intervals and the QT intervals from the 12-lead ECG ( $P=.02$ ). A total of 94% of the average QT interval measurements between the platform and the QT intervals from the 12-lead ECG (lead I) were within the 95% CI for the average difference between the two measurements, with a mean difference of -21.08 (95% LoA -80.34, 38.18). The Bland-Altman analysis detected no significant proportional bias between the AccurBeat QT intervals and the QT intervals from the ECG lead I ( $P=.12$ ). A total of 90% of the average QT interval measurements between the platform and the 12-lead ECG (lead II) were within the 95% confidence limit of the average difference between the two measurements, with a mean difference of -18.48 (95% LoA -78.44, 41.47). The Bland-Altman analysis detected no significant proportional bias between the AccurBeat QT interval and the QT from the ECG lead II ( $P=.51$ ).

Figure 1. The 12-lead corrected QT (QTc) with AccurBeat QTc.

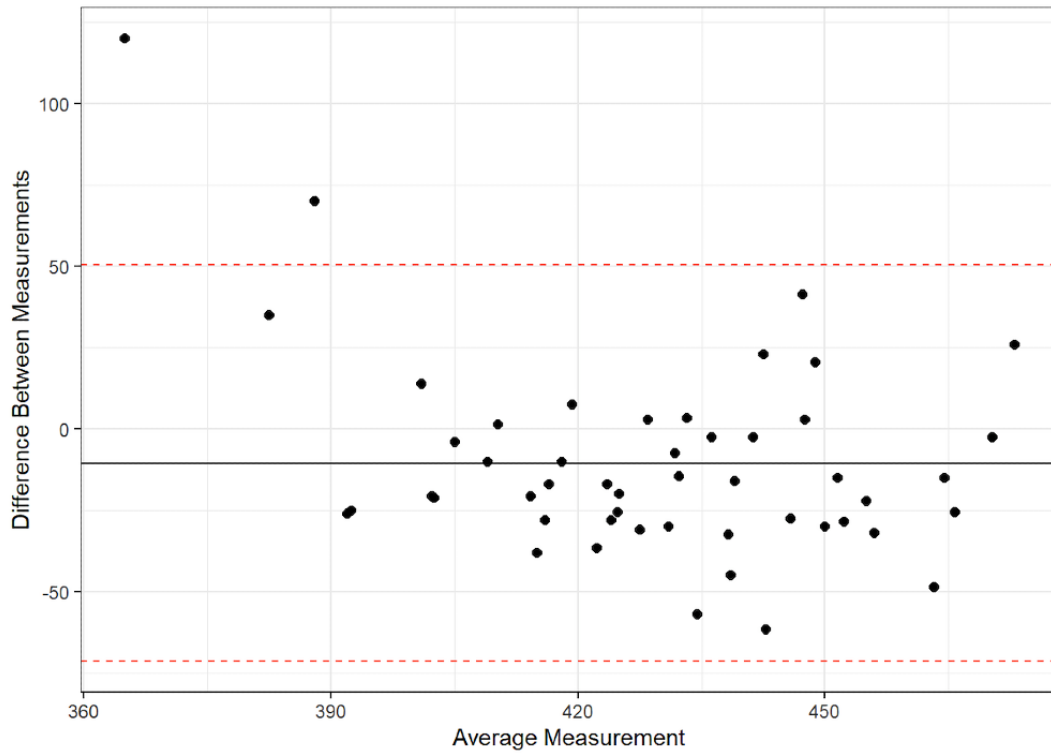


Figure 2. Lead I corrected QT (QTc) with AccurBeat QTc.

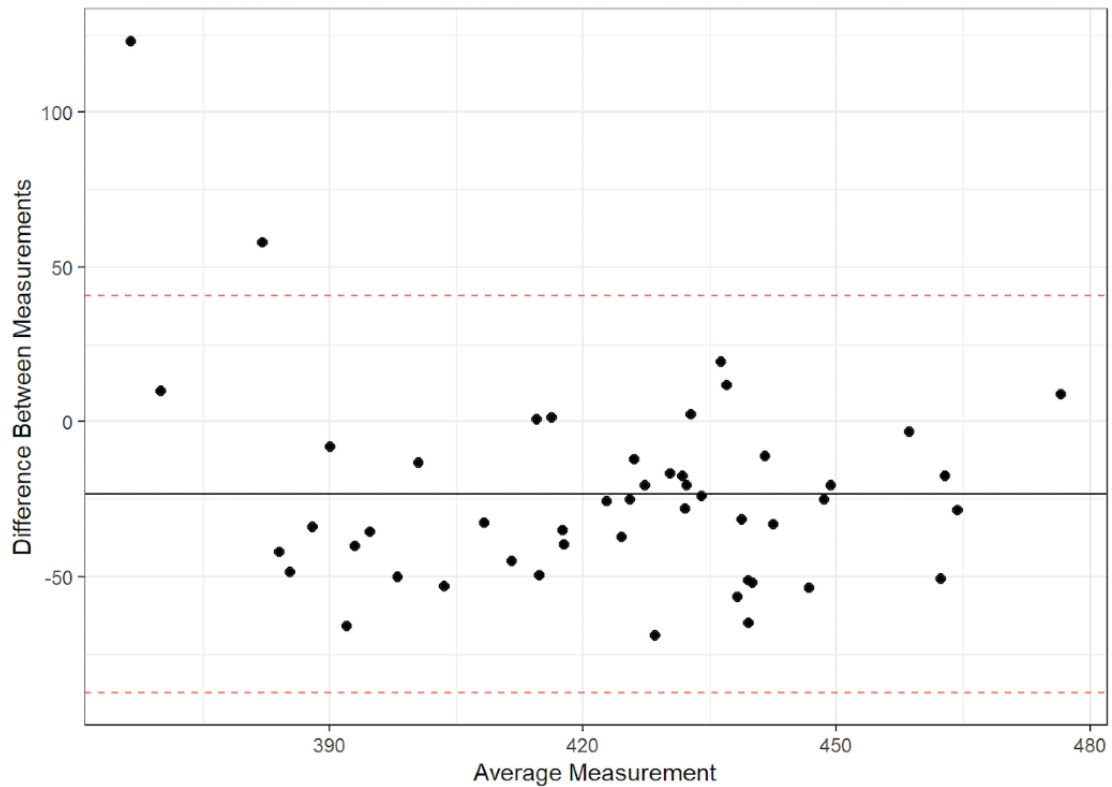


Figure 3. Lead II corrected QT (QTc) with AccurBeat QTc.

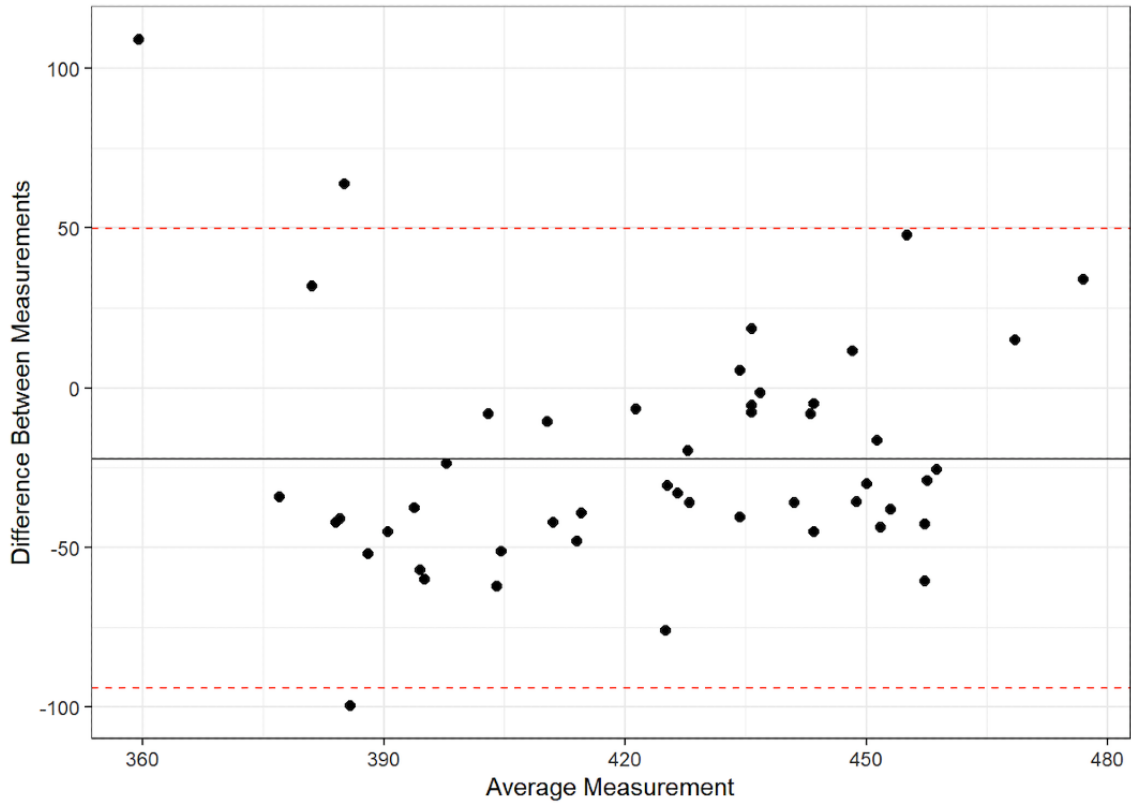


Figure 4. The 12-lead QT with AccurBeat QT.

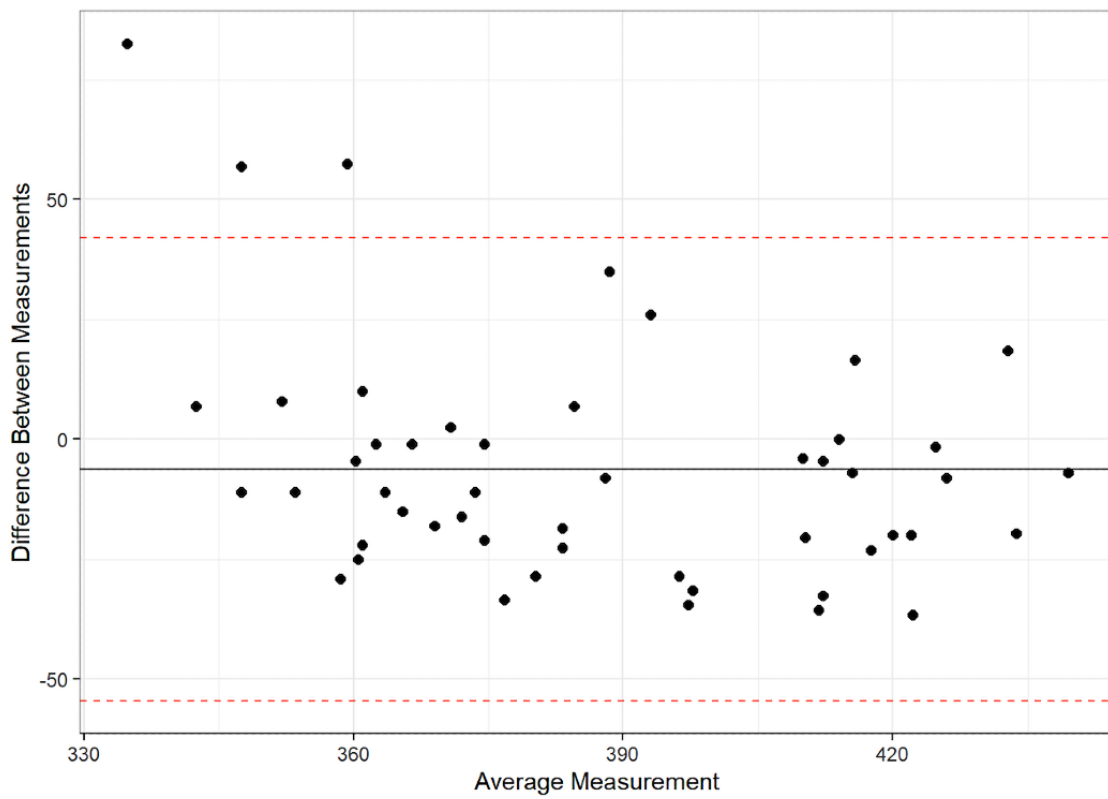


Figure 5. Lead I QT with AccurBeat QT.

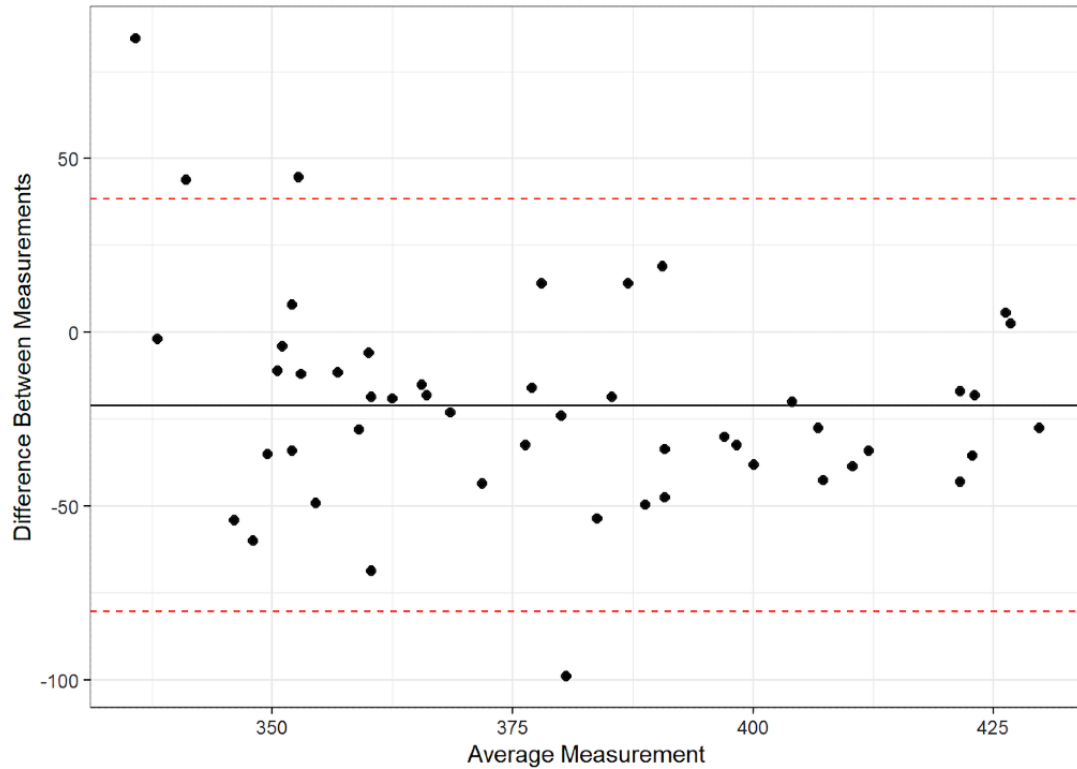
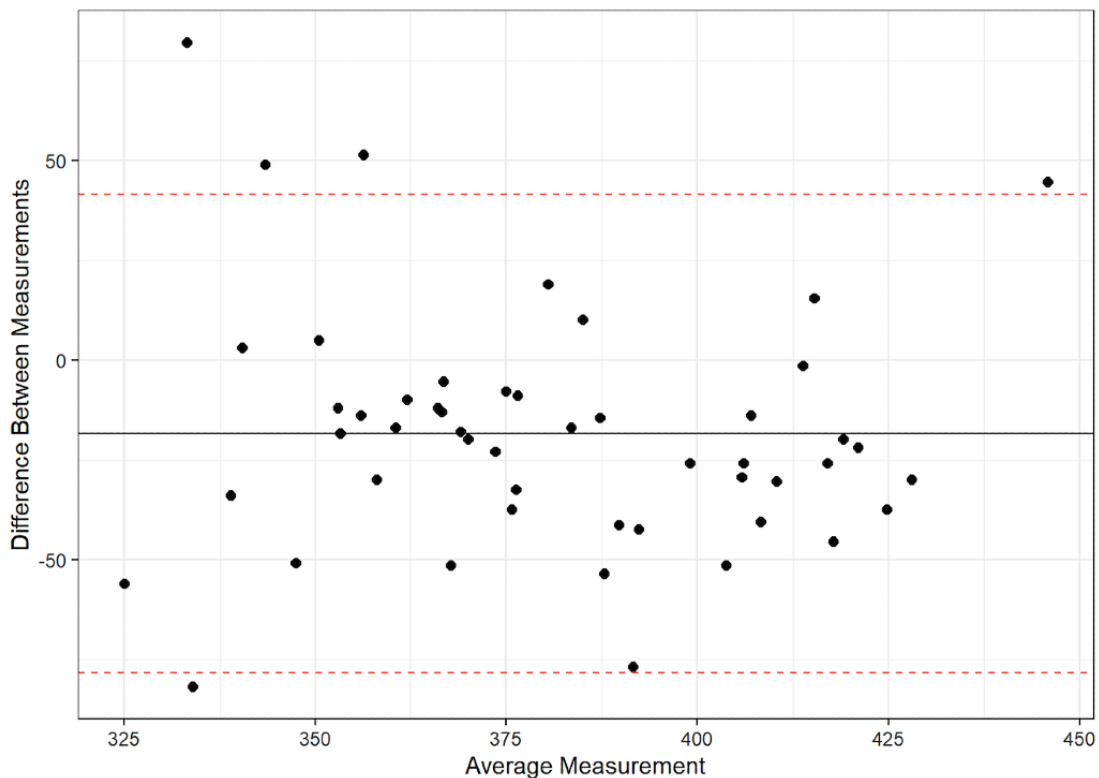


Figure 6. Lead II QT with AccurBeat QT.



## Discussion

### Principal Findings

This study used the Apple smartwatch coupled with the platform to assess QT and QTc intervals, and showed reasonable accuracy with measures derived from conventional 12-lead ECG tracing

in healthy controls without known cardiovascular disease. While associations ranged from low to moderate-high for the various measures of comparison, more than 90% of the average QT interval measurements between the platform and QT intervals from the 12-lead ECG were within the 95% CI of the average difference between the two measurements. Additionally, the technology platform posed no bias in terms of under- or

overestimation. Moreover, intraclass measures of consistency and agreement as well as Pearson correlations were all higher for the QT interval than for the QTc interval.

### Comparison to Prior Work

The measurement of QT intervals is an important consideration in the identification of individuals at increased risk for ventricular tachycardia and sudden cardiac death [8-10]. QT interval monitoring is also important in terms of monitoring patients initiated and dose titrated on various classes of medications. The QT interval represents the time interval from onset of ventricular depolarization to the end of depolarization, is measured from the start of the q wave to the end of the T wave, and is usually obtained from a 12-lead ECG [8]. While computerized automatically derived values are often used for clinical purposes, some authors have advocated that manual measurement is more accurate [24]. However, manual QT interval assessment is tedious and time-consuming with greater interobserver variability, and physicians often select one complex from one lead (lead II) and one to measure [25]. Both techniques are limited by difficult T wave morphologies and the presence of u waves [26].

In recent years, there has been growing interest in the utility of wearables and smartphones [27] for remote ECG monitoring, which has been largely accelerated by the COVID-19 pandemic. In this regard, the Apple Watch has been shown to be a useful screening strategy for the detection of atrial fibrillation [28]. The measurement of QT intervals using smartwatch technology represents an important extension of remote monitoring and poses advantages regarding cost and convenience, and the ability of prolonged monitoring. This study showed comparable values from the Apple smartwatch coupled with the platform technology to those from a 12-lead ECG, from lead I alone and from lead II.

In the multivariate analysis, age did not impact the predictive value of AccurKardia's technology even though age-related changes in QT interval have been previously reported [29]. Although there was strong agreement and consistency for HR values, the comparisons were less strong for QTc intervals than for QT intervals. These findings are not unexpected as QTc interval comparisons include potential measurement errors from two values (QT interval and cycle length), and the QTc interval was calculated by Bazett's [20] formula, in which small differences in HR translate into relatively large differences in QT interval correction. Additionally, while consistency and agreement were similar for most comparisons, agreement was lower in comparisons of smartwatch-obtained values with leads I and II. This finding was expected as QT intervals were measured from 2 to 4 beats in these leads.

Important differences in acquisition methods that could also contribute to potential sources of error include the fact that the

platform's method includes a total of 60 seconds of recording, whereas a 12-lead ECG is recorded over 10 seconds. Since these participants were relatively young (mean age 46 years) and healthy, respiratory variations in HR due to sinus arrhythmia over the short-term recording of a 12-lead ECG could adversely affect comparisons [30]. It may be that 60 seconds of focused application of the Apple Watch was subject to less sinus arrhythmia than 10 seconds of a resting ECG [31]. It remains unclear as to whether comparisons would be even stronger if a longer 12-lead recording time were to be performed. Additionally, the inclusion of patients with heart disease would likely reduce beats due to variations in HR and would expand the range of QT intervals to include prolonged values; in this regard, the design of the study was a conservative one. Typically, there is variability in values obtained from the 12 different leads of an ECG, a phenomenon known as QT dispersion [32,33]. Therefore, the algorithm measured the QT interval from one specific lead as an approximation of the longest QT interval of the 12 leads.

### Limitations

There are a number of limitations in this study that warrant mention. First, smartwatch and 12-lead ECG recordings were obtained sequentially and not simultaneously. However, the brief intervening period should not be expected to substantially impact results. Second, this pilot study evaluated healthy individuals without known heart disease, and the results cannot be extrapolated to those with known heart disease. The research group intends to perform a follow-up study including patients with heart disease. Third, the 12-lead ECG measurements were not made by multiple people but by a single experienced investigator (JL). In future studies, the study group plans to get the 12-lead ECG measured by more than one investigator to take into account interobserver variability.

### Strengths

A strength of the study is that it included a racially diverse group of study participants in which 50% were African American. While darker skin color has been reported to adversely affect smartwatch recording capabilities [34], we did not perform racial comparisons.

### Conclusion

Despite these limitations, we conclude that QT and QTc intervals obtained by the Apple smartwatch coupled with the platform are comparable to those from a 12-lead ECG. Future research is planned to build upon the learnings in this study, expanding to a larger sample of patients that includes patients with cardiac disease as well as multiple cardiologists and assessments of their reliability with regard to manual annotation of ECG tracings. With these learnings and further refinements, remote monitoring using this technology holds promise for the identification of QT prolongation.

### Acknowledgments

NR is also the chief medical officer for AccurKardia. VY, SC, and SK declare that funding for this study was provided to HITLAB by AccurKardia, but the comparative analyses were performed and written independently by HITLAB, Division of Cardiovascular

Medicine, State University of New York Downstate Medical Center, and Department of Epidemiology and Biostatistics, State University of New York Downstate Health Sciences University.

## Data Availability

The data sets used or analyzed during this study will be available from the corresponding author upon reasonable request.

## Conflicts of Interest

SC, RC, SK, JL, GT, and VY report no financial or other relationship relevant to the subject of this paper. NR is the chief medical officer of AccurKardia.

## References

1. 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):e139-e596 [FREE Full text] [doi: [10.1161/CIR.0000000000000757](https://doi.org/10.1161/CIR.0000000000000757)] [Medline: [31992061](https://pubmed.ncbi.nlm.nih.gov/31992061/)]
2. Fryar CD, Chen T, Li X. Prevalence of uncontrolled risk factors for cardiovascular disease: United States, 1999-2010. *NCHS Data Brief* 2012 Aug(103):1-8 [FREE Full text] [Medline: [23101933](https://pubmed.ncbi.nlm.nih.gov/23101933/)]
3. Mant J, Fitzmaurice DA, Hobbs FDR, Jowett S, Murray ET, Holder R, et al. Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial. *BMJ* 2007 Aug 25;335(7616):380 [FREE Full text] [doi: [10.1136/bmj.39227.551713.AE](https://doi.org/10.1136/bmj.39227.551713.AE)] [Medline: [17604299](https://pubmed.ncbi.nlm.nih.gov/17604299/)]
4. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun* 2020 Apr 09;11(1):1760. [doi: [10.1038/s41467-020-15432-4](https://doi.org/10.1038/s41467-020-15432-4)] [Medline: [32273514](https://pubmed.ncbi.nlm.nih.gov/32273514/)]
5. Veronese G, Germini F, Ingrassia S, Cutuli O, Donati V, Bonacchini L, Italian Society of Emergency Medicine (SIMEU). Emergency physician accuracy in interpreting electrocardiograms with potential ST-segment elevation myocardial infarction: Is it enough? *Acute Card Care* 2016 Mar;18(1):7-10. [doi: [10.1080/17482941.2016.1234058](https://doi.org/10.1080/17482941.2016.1234058)] [Medline: [27759433](https://pubmed.ncbi.nlm.nih.gov/27759433/)]
6. Porta-Sánchez A, Gilbert C, Spears D, Amir E, Chan J, Nanthakumar K, et al. Incidence, diagnosis, and management of QT prolongation induced by cancer therapies: a systematic review. *J Am Heart Assoc* 2017 Dec 07;6(12):e007724 [FREE Full text] [doi: [10.1161/JAHA.117.007724](https://doi.org/10.1161/JAHA.117.007724)] [Medline: [29217664](https://pubmed.ncbi.nlm.nih.gov/29217664/)]
7. Nassar Junior AP. COVID-19 pandemic and the opportunity to accelerate remote monitoring of patients. *J Bras Pneumol* 2021 Sep 06;47(4):e20210238 [FREE Full text] [doi: [10.36416/1806-3756/e20210238](https://doi.org/10.36416/1806-3756/e20210238)] [Medline: [34495180](https://pubmed.ncbi.nlm.nih.gov/34495180/)]
8. Postema P, Wilde A. The measurement of the QT interval. *Curr Cardiol Rev* 2014 Aug;10(3):287-294 [FREE Full text] [doi: [10.2174/1573403x10666140514103612](https://doi.org/10.2174/1573403x10666140514103612)] [Medline: [24827793](https://pubmed.ncbi.nlm.nih.gov/24827793/)]
9. Lepeschkin E, Surawicz B. The measurement of the Q-T interval of the electrocardiogram. *Circulation* 1952 Sep;6(3):378-388. [doi: [10.1161/01.cir.6.3.378](https://doi.org/10.1161/01.cir.6.3.378)] [Medline: [14954534](https://pubmed.ncbi.nlm.nih.gov/14954534/)]
10. Giudicessi JR, Schram M, Bos JM, Galloway CD, Shreibati JB, Johnson PW, et al. Artificial intelligence-enabled assessment of the heart rate corrected QT interval using a mobile electrocardiogram device. *Circulation* 2021 Mar 30;143(13):1274-1286 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.120.050231](https://doi.org/10.1161/CIRCULATIONAHA.120.050231)] [Medline: [33517677](https://pubmed.ncbi.nlm.nih.gov/33517677/)]
11. Curran M, Splawski I, Timothy K, Vincent GM, Green E, Keating M. A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell* 1995 Mar 10;80(5):795-803 [FREE Full text] [doi: [10.1016/0092-8674\(95\)90358-5](https://doi.org/10.1016/0092-8674(95)90358-5)] [Medline: [7889573](https://pubmed.ncbi.nlm.nih.gov/7889573/)]
12. Wang Q, Shen J, Splawski I, Atkinson D, Li Z, Robinson JL, et al. SCN5A mutations associated with an inherited cardiac arrhythmia, long QT syndrome. *Cell* 1995 Mar 10;80(5):805-811 [FREE Full text] [doi: [10.1016/0092-8674\(95\)90359-3](https://doi.org/10.1016/0092-8674(95)90359-3)] [Medline: [7889574](https://pubmed.ncbi.nlm.nih.gov/7889574/)]
13. Gaita F, Giustetto C, Bianchi F, Wolpert C, Schimpf R, Riccardi R, et al. Short QT syndrome. *Circulation* 2003 Aug 26;108(8):965-970. [doi: [10.1161/01.cir.0000085071.28695.c4](https://doi.org/10.1161/01.cir.0000085071.28695.c4)]
14. Duncker D, Ding WY, Etheridge S, Noseworthy PA, Veltmann C, Yao X, et al. Smart wearables for cardiac monitoring-real-world use beyond atrial fibrillation. *Sensors (Basel)* 2021 Apr 05;21(7):2539 [FREE Full text] [doi: [10.3390/s21072539](https://doi.org/10.3390/s21072539)] [Medline: [33916371](https://pubmed.ncbi.nlm.nih.gov/33916371/)]
15. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. *JMIR Mhealth Uhealth* 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: [10.2196/12861](https://doi.org/10.2196/12861)] [Medline: [31512582](https://pubmed.ncbi.nlm.nih.gov/31512582/)]
16. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019 Jan;25(1):65-69 [FREE Full text] [doi: [10.1038/s41591-018-0268-3](https://doi.org/10.1038/s41591-018-0268-3)] [Medline: [30617320](https://pubmed.ncbi.nlm.nih.gov/30617320/)]

17. Strik M, Caillol T, Ramirez FD, Abu-Alrub S, Marchand H, Welte N, et al. Validating QT-interval measurement using the Apple Watch ECG to enable remote monitoring during the COVID-19 pandemic. *Circulation* 2020 Jul 28;142(4):416-418 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.120.048253](https://doi.org/10.1161/CIRCULATIONAHA.120.048253)] [Medline: [32478565](https://pubmed.ncbi.nlm.nih.gov/32478565/)]
18. Maille B, Wilkin M, Million M, Rességuier N, Franceschi F, Koutbi-Franceschi L, et al. Smartwatch electrocardiogram and artificial intelligence for assessing cardiac-rhythm safety of drug therapy in the COVID-19 pandemic. The QT-logs study. *Int J Cardiol* 2021 May 15;331:333-339 [FREE Full text] [doi: [10.1016/j.ijcard.2021.01.002](https://doi.org/10.1016/j.ijcard.2021.01.002)] [Medline: [33524462](https://pubmed.ncbi.nlm.nih.gov/33524462/)]
19. Spaccarotella CAM, Migliarino S, Mongiardo A, Sabatino J, Santaripa G, De Rosa S, et al. Measurement of the QT interval using the Apple Watch. *Sci Rep* 2021 May 24;11(1):10817. [doi: [10.1038/s41598-021-89199-z](https://doi.org/10.1038/s41598-021-89199-z)] [Medline: [34031432](https://pubmed.ncbi.nlm.nih.gov/34031432/)]
20. Bazett HC. An analysis of the time-relations of electrocardiograms. *Ann Noninvasive Electrocardiol* 1997 Apr;2(2):177-194. [doi: [10.1111/j.1542-474x.1997.tb00325.x](https://doi.org/10.1111/j.1542-474x.1997.tb00325.x)]
21. 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](https://pubmed.ncbi.nlm.nih.gov/2868172/)]
22. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)* 2015;25(2):141-151 [FREE Full text] [doi: [10.11613/BM.2015.015](https://doi.org/10.11613/BM.2015.015)] [Medline: [26110027](https://pubmed.ncbi.nlm.nih.gov/26110027/)]
23. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979 Mar;86(2):420-428. [doi: [10.1037//0033-2909.86.2.420](https://doi.org/10.1037//0033-2909.86.2.420)] [Medline: [18839484](https://pubmed.ncbi.nlm.nih.gov/18839484/)]
24. Indraratna P, Tardo D, Delves M, Szirt R, Ng B. Measurement and management of QT interval prolongation for general physicians. *J Gen Intern Med* 2020 Mar;35(3):865-873 [FREE Full text] [doi: [10.1007/s11606-019-05477-7](https://doi.org/10.1007/s11606-019-05477-7)] [Medline: [31654357](https://pubmed.ncbi.nlm.nih.gov/31654357/)]
25. Panicker GK, Karnad DR, Natekar M, Kothari S, Narula D, Lokhandwala Y. Intra- and interreader variability in QT interval measurement by tangent and threshold methods in a central electrocardiogram laboratory. *J Electrocardiol* 2009;42(4):348-352. [doi: [10.1016/j.jelectrocard.2009.01.003](https://doi.org/10.1016/j.jelectrocard.2009.01.003)] [Medline: [19261293](https://pubmed.ncbi.nlm.nih.gov/19261293/)]
26. Tran H, Fan C, Tu W, Kertland H, Li L, Kluger J, et al. QT measurement: a comparison of three simple methods. *Ann Noninv Electrocard* 1998 Jul;3(3):228-231. [doi: [10.1111/j.1542-474x.1998.tb00347.x](https://doi.org/10.1111/j.1542-474x.1998.tb00347.x)]
27. Beers L, van Adrichem LP, Himmelreich J, Karregat EPM, de Jong JSSG, Postema PG, et al. Manual QT interval measurement with a smartphone-operated single-lead ECG versus 12-lead ECG: a within-patient diagnostic validation study in primary care. *BMJ Open* 2021 Nov 03;11(11):e055072 [FREE Full text] [doi: [10.1136/bmjopen-2021-055072](https://doi.org/10.1136/bmjopen-2021-055072)] [Medline: [34732504](https://pubmed.ncbi.nlm.nih.gov/34732504/)]
28. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Apple Heart Study Investigators. Large-scale assessment of a Smartwatch to identify atrial fibrillation. *N Engl J Med* 2019 Nov 14;381(20):1909-1917 [FREE Full text] [doi: [10.1056/NEJMoa1901183](https://doi.org/10.1056/NEJMoa1901183)] [Medline: [31722151](https://pubmed.ncbi.nlm.nih.gov/31722151/)]
29. Vink AS, Clur SB, Wilde AA, Blom NA. Effect of age and gender on the QTc-interval in healthy individuals and patients with long-QT syndrome. *Trends Cardiovasc Med* 2018 Jan;28(1):64-75. [doi: [10.1016/j.tcm.2017.07.012](https://doi.org/10.1016/j.tcm.2017.07.012)] [Medline: [28869094](https://pubmed.ncbi.nlm.nih.gov/28869094/)]
30. Soos MP, McComb D. Sinus Arrhythmia. Treasure Island, FL: StatPearls Publishing; 2022.
31. Hatfield BD, Spalding TW, Santa Maria DL, Porges SW, Potts JT, Byrne EA, et al. Respiratory sinus arrhythmia during exercise in aerobically trained and untrained men. *Med Sci Sports Exerc* 1998 Feb;30(2):206-214. [doi: [10.1097/00005768-199802000-00006](https://doi.org/10.1097/00005768-199802000-00006)] [Medline: [9502347](https://pubmed.ncbi.nlm.nih.gov/9502347/)]
32. Hnatkova K, Malik M. Sources of QTc variability: implications for effective ECG monitoring in clinical practice. *Ann Noninvasive Electrocardiol* 2020 Mar;25(2):e12730 [FREE Full text] [doi: [10.1111/anec.12730](https://doi.org/10.1111/anec.12730)] [Medline: [31760674](https://pubmed.ncbi.nlm.nih.gov/31760674/)]
33. Statters D, Malik M, Ward D, Camm A. QT dispersion: problems of methodology and clinical significance. *J Cardiovasc Electrophysiol* 1994 Aug;5(8):672-685. [doi: [10.1111/j.1540-8167.1994.tb01190.x](https://doi.org/10.1111/j.1540-8167.1994.tb01190.x)] [Medline: [7804520](https://pubmed.ncbi.nlm.nih.gov/7804520/)]
34. Ray I, Liaqat D, Gabel M, de Lara E. Skin tone, confidence, and data quality of heart rate sensing in WearOS smartwatches. In: IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events. 2021 Presented at: PerCom Workshops; March 22-26, 2021; Kassel, Germany p. 213-219. [doi: [10.1109/percomworkshops51409.2021.9431120](https://doi.org/10.1109/percomworkshops51409.2021.9431120)]

## Abbreviations

- AI:** artificial intelligence
- ECG:** electrocardiogram
- HR:** heart rate
- LoA:** limits of agreement
- QTc:** corrected QT
- SUNY:** State University of New York
- VIF:** variance inflation factor



*Edited by A Mavragani; submitted 20.07.22; peer-reviewed by R Widmer, J Piera-Jiménez; comments to author 16.08.22; revised version received 12.09.22; accepted 16.09.22; published 28.09.22.*

*Please cite as:*

*Chokshi S, Tologonova G, Calixte R, Yadav V, Razvi N, Lazar J, Kachnowski S  
Comparison Between QT and Corrected QT Interval Assessment by an Apple Watch With the AccurBeat Platform and by a 12-Lead  
Electrocardiogram With Manual Annotation: Prospective Observational Study*

*JMIR Form Res 2022;6(9):e41241*

*URL: <https://formative.jmir.org/2022/9/e41241>*

*doi: [10.2196/41241](https://doi.org/10.2196/41241)*

*PMID: [36169999](https://pubmed.ncbi.nlm.nih.gov/36169999/)*

©Sara Chokshi, Gulzhan Tologonova, Rose Calixte, Vandana Yadav, Naveed Razvi, Jason Lazar, Stan Kachnowski. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 28.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Thinking Aloud or Screaming Inside: Exploratory Study of Sentiment Around Work

Marzia Hoque Tania<sup>1</sup>, PhD; Md Razon Hossain<sup>2</sup>, MSc; Nuzhat Jahanara<sup>3</sup>, MS; Ilya Andreev<sup>4</sup>, BSc; David A Clifton<sup>1,5</sup>, DPhil

<sup>1</sup>Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom

<sup>2</sup>School of Information System, Queensland University of Technology, Brisbane, Australia

<sup>3</sup>Department of Psychology, University of Dhaka, Dhaka, Bangladesh

<sup>4</sup>School of Engineering and the Built Environment, Anglia Ruskin University, Cambridge, United Kingdom

<sup>5</sup>Oxford Centre for Advanced Research (OSCAR), University of Oxford, Suzhou, China

**Corresponding Author:**

Marzia Hoque Tania, PhD

Institute of Biomedical Engineering

Department of Engineering Science

University of Oxford

CHI Lab

Old Road Research Campus

Oxford, OX3 7DQ

United Kingdom

Phone: 44 1865617670

Email: [marzia.hoquetania@eng.ox.ac.uk](mailto:marzia.hoquetania@eng.ox.ac.uk)

## Abstract

**Background:** Millions of workers experience work-related ill health every year. The loss of working days often accounts for poor well-being because of discomfort and stress caused by the workplace. The ongoing pandemic and postpandemic shift in socioeconomic and work culture can continue to contribute to adverse work-related sentiments. Critically investigating state-of-the-art technologies, this study identifies the research gaps in recognizing workers' need for well-being support, and we aspire to understand how such evidence can be collected to transform the workforce and workplace.

**Objective:** Building on recent advances in sentiment analysis, this study aims to closely examine the potential of social media as a tool to assess workers' emotions toward the workplace.

**Methods:** This study collected a large Twitter data set comprising both pandemic and prepandemic tweets facilitated through a human-in-the-loop approach in combination with unsupervised learning and meta-heuristic optimization algorithms. The raw data preprocessed through natural language processing techniques were assessed using a generative statistical model and a lexicon-assisted rule-based model, mapping lexical features to emotion intensities. This study also assigned human annotations and performed work-related sentiment analysis.

**Results:** A mixed methods approach, including topic modeling using latent Dirichlet allocation, identified the top topics from the corpus to understand how Twitter users engage with discussions on work-related sentiments. The sorted aspects were portrayed through overlapped clusters and low intertopic distances. However, further analysis comprising the Valence Aware Dictionary for Sentiment Reasoner suggested a smaller number of negative polarities among diverse subjects. By contrast, the human-annotated data set created for this study contained more negative sentiments. In this study, sentimental juxtaposition revealed through the labeled data set was supported by the n-gram analysis as well.

**Conclusions:** The developed data set demonstrates that work-related sentiments are projected onto social media, which offers an opportunity to better support workers. The infrastructure of the workplace, the nature of the work, the culture within the industry and the particular organization, employers, colleagues, person-specific habits, and upbringing all play a part in the health and well-being of any working adult who contributes to the productivity of the organization. Therefore, understanding the origin and influence of the complex underlying factors both qualitatively and quantitatively can inform the next generation of workplaces to drive positive change by relying on empirically grounded evidence. Therefore, this study outlines a comprehensive approach to capture deeper insights into work-related health.

**KEYWORDS**

work-related mental health; sentiment analysis; natural language processing; occupational health; Bayesian inference; machine learning; artificial intelligence; mobile phone

## *Introduction*

### **Background and Motivation**

The economic growth of the human civilization, along with our understanding of medical science in combination with the recent technological advancements, has brought us to question how to design the workforce management system in the era of Industry 4.0. The image of occupational health has been gradually shifting from its strong association with workplace injuries to work-related ill health. Among the European nations, most of the UK workforce trust that their health or safety is not at risk because of their workplace [1]. However, much remains to be done to achieve more efficiency in workplace management as millions of workers are still experiencing work-related illnesses every day [1,2].

According to the recent Health and Safety Executive (HSE) report covering the period from 2020 to 2021, 1.7 million workers experienced a work-related illness, approximately half of which were due to stress, depression, or anxiety [1]. In the pre-pandemic report, >800,000 workers experienced mental health issues, whether a new presentation or a long-standing condition, because of the workplace, resulting in the loss of 17.9 million working days. According to the Labour Force Survey (LFS) studies, workload, extreme deadlines, excessive responsibility, absence of adequate managerial support, violence, threats or bullying, and changes at work such as reformation are estimated to be the main reasons behind such poor mental health [2]. Although factors vary from industry to industry, in terms of occupation, professional occupations (examples include, but are not limited to, scientists, engineers, programmers, health workers, and teaching and educational professionals) show higher levels of stress compared with all jobs.

If work-related mental health was not already a concern, COVID-19 has added more dimensions to it. The increase in depression among adults in the United Kingdom during the pandemic has been well reflected in the latest Office for National Statistics report [3]. Although the HSE presented self-reported work-related stress, depression, and anxiety from the LFS study, the finding echoes mental health figures among the adult population in the United Kingdom [4]. The HSE report also quantifies how work-related well-being is affected among different age groups and genders. The gender difference in mental disorders such as depression and anxiety is fairly common, as reported by the World Health Organization [5]. Therefore, formulating a resilient workplace would require a comprehensive understanding of the age and gender disparity in the loss of working days to circumvent human as well as algorithmic bias.

By their very nature, mental health issues and well-being are difficult to measure, and the HSE has 2 different data sources

from which analyses are conducted—none of which record real-time data. Moreover, work-related new or long-standing ill health as a consequence of long or irregular working hours, stress, anxiety, panic, hidden or unrealistic expectations, job insecurity, instability in the job market, societal pressure, rat race, and macho culture, which contributes to an unhealthy lifestyle, often gains visibility in the form of eating disorders, irregular or insufficient sleep, and addiction.

To probe beneath symptoms on the surface by separately tagging each attribute with the associated influence as well as contributing factors adds more complexity when exposing such vulnerability may result in job loss or at least an obstruction to career progression. Hence, it is understandable that approximately 15% of working adults show indications of symptoms of an existing mental health condition [6] as >300,000 people each year lose employment because of mental health problems [7]. By contrast, technological advancement has opened the door to harnessing the power of wearable sensors, the Internet of Things, and artificial intelligence (AI) to gather richer insights on general mental health issues among the population as well as industry-specific personalized circumstances (within and because of the workplace).

### **State-of-the-art Technological Support**

Exploring recent quartile 1 and 2 journals and top conference publications using search words such as “mental health AND artificial intelligence,” “mental health AND decision support system,” “mental health AND mHealth,” and “mental health AND mobile apps” through Google Scholar, PubMed, PsycArticles, ScienceDirect, and Psychology and Behavioral Sciences Collection search engines, we discovered that the broader domain in the literature covers the area of severe mental illness such as schizophrenia [8-11]; anxiety disorders [12] such as posttraumatic stress disorder [13]; developmental disorders such as attention-deficit/hyperactivity disorder [14]; and, of course, disorders that are often not recognized as an outcome of occupational health hazards affecting the everyday quality of life [15,16]—however, neglecting these early indications may even trigger suicidal tendencies [17-19].

Similar to physical health, state-of-the-art investigations on mental health aspects are mainly geared toward screening, diagnosing, and phenotyping purposes [20-23]. However, web-based [24,25] and mobile-based interventions [26-29] are not far behind.

A wide range of technology-enabled support has been investigated in the literature, starting from simple tools such as traditional SMS text messages [30] to cutting-edge technology such as AI-enabled chatbots [31]. Mental health and well-being apps have millions of installations worldwide [32], which makes the demand for such apps evident; however, very few studies have been conducted to verify and validate the capabilities of these apps to bring positive changes in users' mental well-being.

With or without smartphone integration, recent advancements in AI have brought some commercial successes, such as Babylon [33], Quarte [34], Lyra [35], Ginger [36], Woebot [37], and BioBeats [38], to support health and well-being, backed up by collaborative research with world-leading universities. Most of these available apps on the market consist of AI-enabled chatbots and use natural language processing (NLP), such as the AI-driven personalized triage and symptom checker tool by K [39] and CBT by Woebot.

Within the mental health area using data-driven approaches, a New York-based start-up, Spring Health [40], has developed an AI-driven personalized triage and symptom checker for the mental health of employees, whereas a London-based company called BioBeats provides an AI-enabled intervention for stress management. BioBeats developed an intelligent app with a business-to-business model that provides a well-being score based on physiological data (sleep duration and quality and heart rate variability), psychological data (mood journaling), and neuropsychological data (brain function tests) collected through their BioBase mobile app and BioBeam wearable. Tools such as BioBeats can enable their users to gain insights into personal mental health. Such scientifically validated well-being tools can be thought of as company-provided perks. By gathering aggregated and real-time but anonymized data from employees using the BioBeats platform, the employer can track the well-being of their staff and take actions such as providing tailored support when needed. Continuous monitoring can help understand and even quantify whether a change in the company's policy is compromising the well-being of the employees.

### **Sentiments on Social Media Platforms: Related Works**

Microblogging today has become a popular communication tool among internet users. Millions of messages appear daily on popular websites that provide services for microblogging, such as Twitter and Facebook. This is due to the nature of microblogs, on which people post real-time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiments about their daily lives. Owing to the free format of the messages and easy accessibility of microblogging platforms, internet users tend to shift from traditional communication tools (such as traditional blogs or mailing lists) to microblogging services. Therefore, the use of social media has become an integral part of daily routine in modern society.

In terms of the choice of microblogging platform, privacy concerns have been observed in the literature [41]. The type of self-disclosure also has an impact on the use of such social media [41]. It has been seen that the users of Twitter have high self-disclosure. People who like to bond socially prefer using Facebook, as suggested by Shane-Simpson et al [41].

Some of the positive use cases of social media include pandemic studies. Twitter has been used to explore diverse issues such as sentiment alteration [42], lockdown [43], sentiment around hospital care management [44], vaccination [45], and remote working [46]. The use of social media for health and well-being research is common, but it is mostly limited to participant recruitment. However, in a study in San Diego, the health

outcomes of the local people were observed using tweets that measured self-rated mental health, sleep quality, and heart disease [47].

Previously, data from social media profiles of US military personnel have been used by Bryan et al [48] to find predictors of suicide. Analysis of the tweets revealed various aspects of their lives along with the triggering factors in stressful situations, such as health issues, maladaptive or avoidant coping strategies, emotional state, and cognition. It has been observed by Bryan et al [48] that the manner of posting can differ between suicidal and nonsuicidal users. A pattern of posts was also observed as the “trigger” posts increased before suicide following negative emotional posts. Suicidal people who posted about maladaptive coping frequently were followed by a few negative emotional posts [48].

Among the working adults in the United Kingdom, suicide is more prevalent in the construction industry [49]; the use of drugs, alcohol, marijuana, and other substances is also common among this workforce [50]. Although not specified for construction workers, using data from Twitter-based advertisements, the reasons for using marijuana and the characteristics of marijuana users were explored in the reported article [51].

Several studies have investigated the psychological rationale for the temptation to share on social media and the mental profile of those who share on social media and to what extent [52]. However, a more critical exploration with clinical validation is required. By contrast, people share on social media to maintain social connections even if it is not a pathological need.

Using adolescent population data from thousands of people aged 14 years, Kelly et al [53] attempted to investigate whether mental health is linked to the use of social media. The study concluded that a positive correlation exists between the use of social media and depressive symptoms. Other factors that influence the use of social media are poor sleep, low self-esteem, poor body image, and harassment on the internet. These factors are the underlying symptoms of depression as well. Social media use can result in increased perceived social isolation [54]. Too much social media use can also cause social media fatigue, affecting psychological well-being [55].

Although the aforementioned studies show the psychological aspect of why people share on social media, the underlying rationale of social media use by individuals, as well as the impact, can be more complex [56]. Moreover, technological tools such as “Gamification” using neurological hacks can influence social media use, as observed by Bell et al [57].

Social media could be an innovative tool for interventions. However, the impact can go either way as Weinstein [58] observed a seesaw effect of social media while exploring variables such as relational interactions, self-expression, interest-driven exploration, and browsing.

Interestingly, the deepest insights within the context are owned by the social media industry, and third parties in the name of “personalization” and “enhanced experience,” which have the ability to trigger social media addiction, are yet to be extensively studied.

The “Big Five personality traits” (ie, neuroticism, agreeableness, conscientiousness, openness, and extroversion [59]) can be traced from the data obtained from social media using digital footprints [60]. Estimation and quantification of personality traits can support the use of social media for the greater good beyond the personalization feature of social media itself. It is important to associate the understanding of the psychological need to “share” emotions on internet-enabled platforms with the machine-learned quantified measurement of sentiments before taking actions based on an under- or overestimation of our emotions.

## Research Aim

This paper presents an investigation into the methods and tools to understand and analyze how sentiments around work are expressed, which can be further used to flag such issues and design effective interventions. This paper presents a literature review in the Introduction section exploring both reported articles and available state-of-the-art technological support. This paper also provides criticism on vulnerable aspects. The key contributions of this study include the following: (1) to the best of our knowledge, this is the first work performing sentiment analysis and topic detection concentrating on work-related sentiments comprising both pandemic and prepandemic tweets; (2) we collected a large data set based on a hybrid approach for keyword search; (3) we created a labeled data set; and (4) the preprocessed complete data set, as well as the labeled data set on work sentiment, will be made available, which may create opportunities for further studies.

The rest of the paper is organized as follows. The Methods section presents the methods used to critically investigate work-related sentiments on Twitter, followed by the analysis

of the results in the Results section. This paper includes an in-depth discussion in the Discussion section highlighting the intertwined vicious cycle of physical and mental health and also the ethical concerns regarding the use of technology. This paper also discusses a conceptual framework for AI-enabled mental health support systems for the workforce. Finally, we discuss the study’s limitations and scope for improvement while concluding the paper.

## Methods

### Microblogging Platform

This study explores work-related sentiments on Twitter. Twitter, with >319 million monthly active users, has now become a goldmine for researchers, organizations, and individuals to survey public health trends because of the nature of the data source. Twitter allows developers to fetch data (ie, tweets) from its archive using an application programming interface. On the basis of the literature, it is anticipated that AI-enabled tools such as sentiment analysis will help us better understand how people talk about and feel with respect to specific health topics or conditions—in this case, mental health—related issues linked to work—in real time.

### Data Fetching and Processing

In this study, we used the Python-based library SNScrape (Scraper for Social Networking Services) [61] to fetch tweets using the defined primary keywords from [Textbox 1](#). We used these keywords to formulate the keywords for fetching tweets according to the procedure described by Edo-Osagie et al [62]. The search query, dated from January 1, 2018, to December 31, 2021, resulted in >1 million tweets containing at least one keyword either as a hashtag or in the tweet body.

**Textbox 1.** Primary keywords used to form the keywords for fetching tweets.

#### Primary keywords

- Nouns: “work,” “job,” “industry,” “labor,” “labor,” “office,” “workplace,” “occupation,” “employment,” “employee,” “employer,” “corporate,” “company,” “enterprise,” “profession,” “meet deadlines,” “cubicle,” “workplace”
- Adjectives: “stress,” “anxiety,” “depression,” “inefficiency,” “multitasking,” “breaking point,” “overload,” “burnout,” “stressed out,” “under pressure,” “demoralization,” “depressed mood,” “horrific,” “shocking,” “dangerous,” “problematic,” “undesirable,” “sad,” “struggle,” “stressed,” “depressed,” “ill,” “illness,” “mental strain,” “best lucky,” “bepositive,” “bethet united,” “get in,” “vibes,” “ftw,” “accomplished,” “let’s get it done,” “mic dropped alhamdulillah,” “agreed,” “happy,” “good,” “enjoyment,” “proud,” “satisfaction,” “bad,” “unlucky,” “sorry,” “bad,” “worse,” “bad,” “hurt,” “tiring,” “long-day,” “boring,” “exhausted,” “hopeful,” “2ndhome,” “standwith,” “godspeed,” “more to follow”
- Phrases: “overwork,” “wage slave,” “ironwork,” “mental health,” “congratulations office,” “work life,” “workoholic,” “worklifebalance,” “job strain,” “bad day at the office,” “when going gets tough workaholic gets going,” “comeback,” “back to drawing board,” “work-life balance”

## Keyword Selection Process

### Hybrid Steps

The Twitter search uses words to find tweets relevant to the quest. These words or parts of sentences are called keywords. Proper keywords help to find and fetch tweets that are efficient and accurate to the context [63]. Therefore, the selection of keywords is a crucial step in fetching tweets. We followed a 2-fold process to select the keywords.

### Phase 1: Keyword Collection From a Panel

First, a panel of working professionals was formed to estimate the linguistic diversity in the expression of emotions regarding the workplace and work. There were 7 members in the panel, and they were proficient users of social media, had a successful background in tertiary education, and were in a responsible position at work. The responses were collected using Google Forms. We received responses from a diverse range of industries and professions, including software engineers, corporate officers, and academics. The panel also involved experts from work rights organizations. The tenure of work experience was also

recorded as background information; however, it was not used in the data analysis because of the size of the panel.

In this phase, the panel members were asked how they articulated work-related sentiments on social media, specifically on Twitter. The responses were recorded anonymously. The panel members were asked about their linguistic preference to denote work and the workplace (inclusive of words and phrases with similar semantic meanings and similar sentiment-expressing adjectives, whether positive, negative, or neutral). We divided these words into 3 categories, which are displayed in [Textbox 2](#).

Category 1, also denoted as nouns in [Textbox 1](#), represents the semantic expression of “work,” manually identified by the natural language users (ie, the panel of working adults on Twitter). These keywords mean only something related to work. In contrast, category 2 aims to capture the core sentiment. The set of words in category 2 can be related to anything, including work and the workplace. However, adding an adjective to the synonyms of *work* and *workplace* means that the sentiment expressed is related to work.

For example, the words “office” or “employer” alone do not always express the sentiment regarding work life and can be used in many other ways, such as the address of a workplace or a job circular.

Moreover, the adjectives “struggle” or “good” alone can also be related to anything. However, to increase the search

efficiency and fetch more contextual tweets, we searched for tweets that contained both “office” and “good” or “office” and “struggle.”

Furthermore, the participating professionals were from a particular region and were limited in number. Although there was diversity in their occupations, the words and phrases they used certainly cannot completely represent or match the words and phrases of millions of other professionals and employees. Therefore, we can obtain more possible forms of keywords by combining each word in category 1 with each word in category 2.

In a nutshell, the reasons behind creating category 1 and category 2 are that (1) to search tweets that contained work-based sentiment, we would need related words or synonyms to “work” or “workplace” (category 1) and words that expressed “sentiment” or “emotion” regarding these (category 2) and (2) to fetch more diverse and the highest number of tweets, we needed to pragmatically maximize all possible forms of keywords and, therefore, separating them into different categories, represented as rows in [Textbox 1](#), helped us combine category 1 and 2 with each other and increase the number of keywords.

Subsequently, we also obtained some phrases from the professionals (category 3) that directly describe work-related sentiments, such as “wage slave” and “congratulations office.”

**Textbox 2.** Categories of words and phrases in [Textbox 1](#) used to denote work and the workplace.

#### Categories of words and phrases in [Textbox 1](#)

- Category 1: words and phrases implying work, the office, or the workplace
- Category 2: words and phrases signifying sentiments (can be positive, negative, or neutral)
- Category 3: phrases that explicitly mean sentiments regarding work, the office, or the workplace

### Phase 2: Keyword Formulation Using Algorithms

Despite the careful choice of a panel of working professionals, there was a need for an enhanced set of keywords owing to human bias, spelling practices on social media, the internet, and work-related jargon. Moreover, we considered only English keywords, whereas there are thousands of languages in the world. Hence, the collected words and phrases can be limited, and there can be far more words and phrases used by millions of workers worldwide. Therefore, to generate more keywords, we used the Global Vectors for Word Representation (GloVe), which process billions of tweets and generate synonyms based on the context. The workflow in this phase is shown in [Textbox 3](#).

GloVe is a well-known algorithm that uses billions of tweets to generate synonyms with a context. Therefore, in the second phase, we used GloVe to generate 5 synonyms for each word in categories 1 and 2 from [Textbox 1](#). We considered all the words in category 1 and their generated synonyms as a prefix list and all the words in category 2 and their generated synonyms

as a postfix list. These prefix and postfix lists contain unique words. For each word in the prefix list, we added all the words from the postfix list and generated the seed keyword list. Let us assume that the prefix list is (workplace, corporate) and the postfix list is (stress, good). Therefore, the seed keyword list is (workplace stress, workplace good, corporate stress, corporate good).

We used each item from the “seed keyword list” to fetch the seed tweets to define the final keywords. For each item on the list, we split the words based on space and searched for tweets that contained these words. For example, the item “workplace stress” in the aforementioned “seed keyword list” contains 2 words: “workplace” and “stress.” Therefore, we searched tweets that contained both “workplace” and “stress” in any sequence and any number.

Subsequently, we used these seed tweets and seed keywords to generate the final keywords using a powerful metaheuristic optimization algorithm (ie, the particle swarm optimization algorithm).

**Textbox 3.** Flow of phase 2.

**Flow**

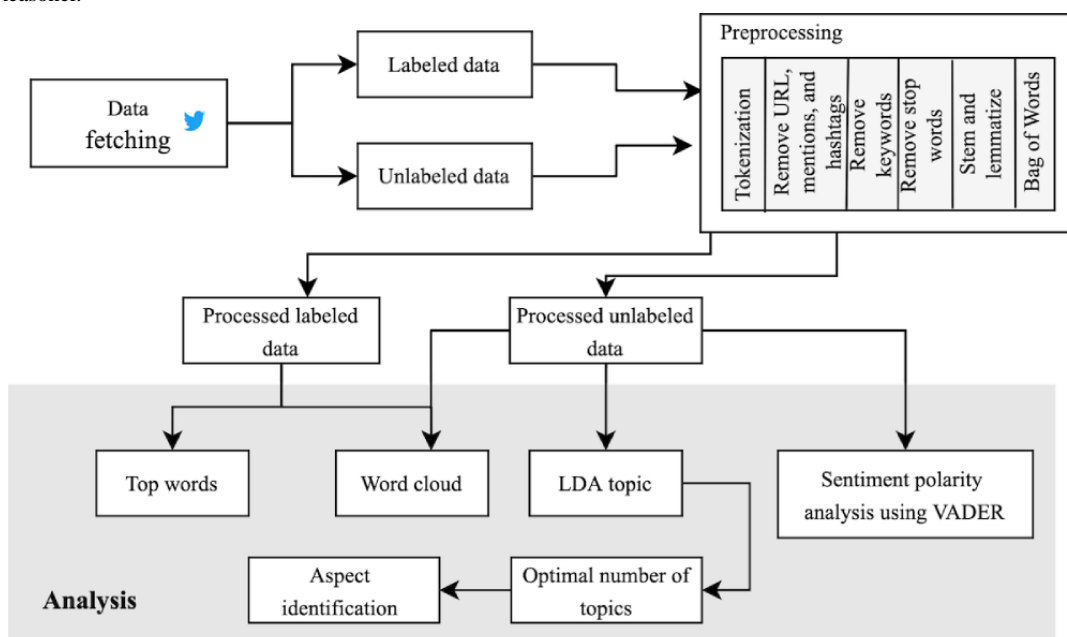
- For each word or phrase in category 1 and category 2, generate 5 synonyms
- Consider all words and phrases in category 1 and their synonyms as a prefix list
- Consider all words and phrases in category 2 and their synonyms as a postfix list
- To generate all possible seed keywords (eg, the keyword list is *KT*):
  - For each word or phrase in the prefix list: *Wpre*
    - For each word or phrase in the postfix list: *Wpst*
      - $KT.add(Wpre+' '+Wpst)$
- Fetch seed tweets by using this *KT*
  - For each keyword in *KT*:
    - If the keyword contains more than one word:
      - Split the words
      - Fetch tweets that contain these words
- Use these seed keywords and fetched seed tweets in the particle swarm optimization algorithm to generate actual keywords

**Data Preprocessing**

The aim of preprocessing is to clear all the redundant and unnecessary content from the data and make them precise to find more accurate words that can help analyze the tweets. With this goal, the fetched tweets were systematically preprocessed (Figure 1) using a number of Python libraries such as Pandas, NumPy, and Natural Language Toolkit (NLTK) before initiating the analysis.

Each tweet was converted into lower case to easily identify the repetitive words and tokenize to convert the tweet into a list of words. The elimination of the selected keywords facilitated a better estimation of the frequency of words and easy spotting of the actual words used in expressing a sentiment. The contractions were expanded to their original forms. The data cleaning steps also included the removal of alphanumeric words, URLs, markup texts, mention words, stop words, hashtags, repeated characters, punctuation symbols, white spaces, and single characters from the tweets. Subsequently, normalization was achieved through lemmatization and stemming.

**Figure 1.** Preprocessing and analysis steps for both labeled and unlabeled data. LDA: latent Dirichlet allocation; VADER: Valence Aware Dictionary for Sentiment Reasoner.



The purpose of this study was to inspect personal tweets, not organizational or commercial tweets. In an attempt to separate such tweets, more words were required to be eliminated, which are marked as advertisement words in [Textbox 4](#). In addition, for simplicity of analysis, incomplete and misspelled words and unrecognizable shortened words had to be removed as well.

**Textbox 4.** Advertising and meaningless words in the labeled data set.

#### Meaningless and advertising words

- Meaningless words: *thi, wa, WA, rt, and ann*
- Advertising words: *click and coupon*

## Data Processing and Labeling

This paper presents an investigation into the tweets ([Textbox 1](#)) through the latent Dirichlet allocation (LDA) [64] model, using the Gensim (RARE Technologies Ltd) [65] library to explore issues discussed on Twitter and the Valence Aware Dictionary for Sentiment Reasoner (VADER) [66] model to estimate the sentiment of each tweet.

The investigation began with the identification of top words within the processed data set. The corpora of Gensim returned an ID term dictionary object, which subsequently created a corpus by converting each tokenized tweet into a word matrix. Later, the LDA model-generating function from the Gensim library used this dictionary and corpus and returned an LDA Model. This LDA Model consisted of a number of specific topics. Each topic contained a number of words along with their weight in the tweet. Using LDA Model, all the tweets were divided into several topics, which enabled us to understand the aspect of each tweet that can be representative of a broader classification.

However, overlap among these topics was expected, which might hinder the identification of an aspect of a topic accurately. Hence, it was necessary to explore the optimal number of topics. Each LDA Model with a specific number of topics had a specific coherence. Coherence measures the score of semantic similarity between the contexts of multiple documents and, the higher the coherence value, the better the context-based topic distribution [67]. Therefore, the LDA Model with higher coherence has an optimal number of topics and, consequently, fewer overlapped topics. Moreover, the LDA Mallet wrapper on top of LDA provides better topic distribution and coherence and, therefore, we used the LDA Mallet wrapper to identify the optimal number of topics [68]. Afterward, we converted this optimized number of LDA Mallet topics into LDA Model topics for visualization, as shown in [Figure 2](#). Such optimization was achieved through the iterative process of varying topics from 2 to 31. The coherence value for each iteration is shown in [Multimedia Appendix 1](#).

Consequent to topic optimization, broader polarity of sentiments (ie, positive, negative, and neutral) was assigned with the help

The quality of the preprocessing step was verified through a qualitative inspection of random processed tweets. This inspection was carried out after each preprocessing step. The qualitative inspection also supported the recognition of meaningless words and advertising tweets.

of the opinion dictionary library from NLTK. The aspect exploration process used in this study can also be observed in algorithm 1 ([Textbox 5](#)).

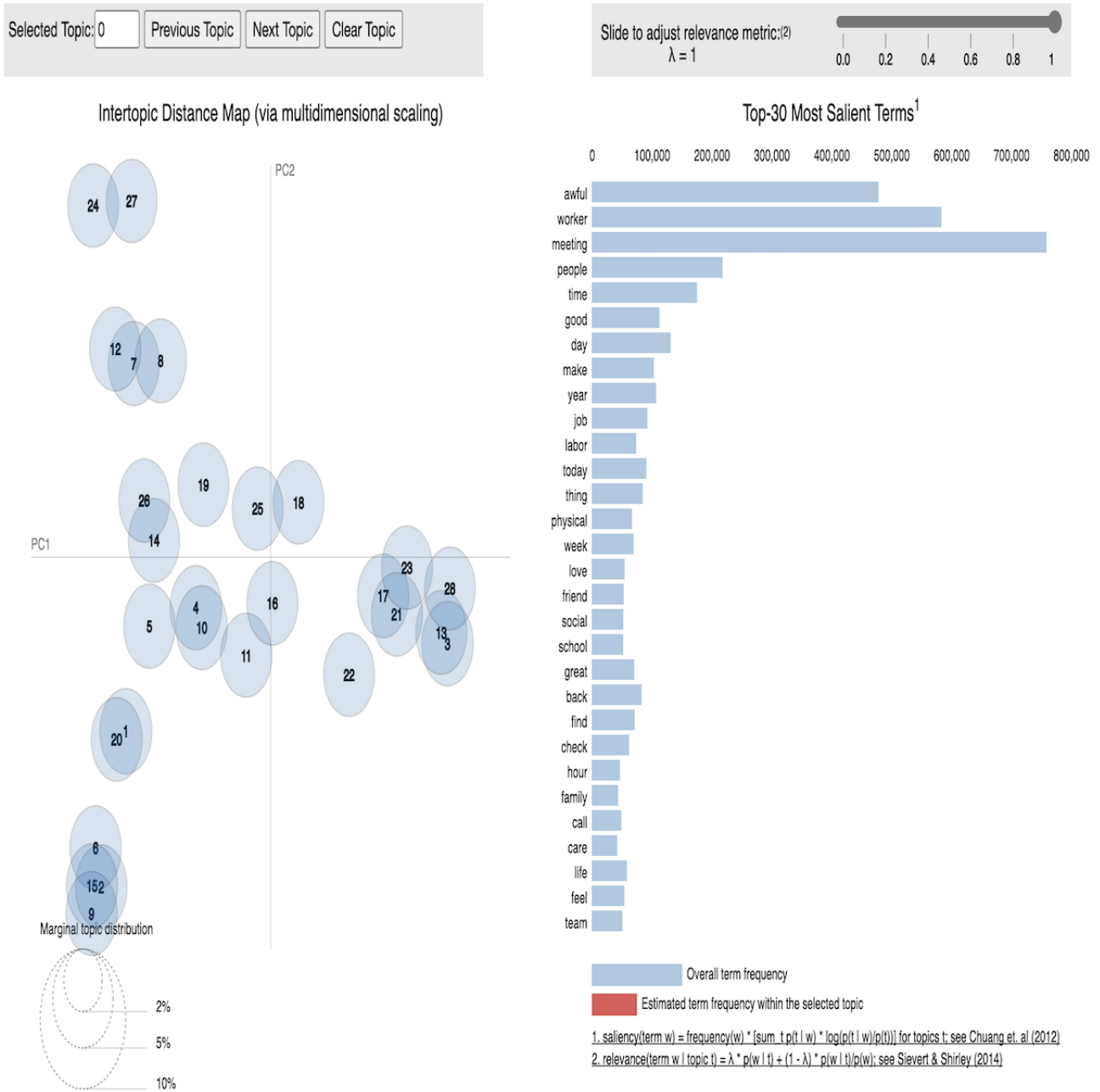
Next, a lexicon and rule-based model, VADER, was used to categorize the unlabeled tweets into our predefined 3 categories (positive, negative, and neutral) of sentiment polarity by analyzing each tweet. VADER considered the sentiment lexicon used in social media microblogs and generalized grammatical and syntactical aspect rules for identifying sentiment intensity. Moreover, it incorporated a human-centric approach combining qualitative analysis with empirical validation and experimental investigations, leveraging the wisdom of the crowd to calculate values based on the sentiment of a sentence. Furthermore, the performance of their algorithm with a lean lexicon and rule-based model was compared with other well-known and familiar sentiment analysis benchmarks [66], estimating a compound value ranging from  $-0.05$  to  $+0.05$  by considering the positivity and negativity of the tweet. If the sentiment value of the tweet was between  $-0.05$  and  $+0.05$ , the tweet was neutral. A tweet with a value  $<-0.05$  was negative, and a tweet with a value  $>+0.05$  was considered positive. However, the lower the value, the higher the negative intensity, and the higher the value, the higher the positive intensity.

Reckoning the compound values and visualization, the percentage of each category aided us in determining which category had a greater number of tweets, whereas visualization of the compound values of each tweet helped us realize the fluctuations in values for each polarity.

In addition to the machine-assigned polarity of sentiments, this study included manual labeling of a subset of 3200 unprocessed random tweets as a pragmatic approach to further analyze positive, negative, and neutral tweets, with a higher chance to recognize juxtaposition of sentiments and sarcasm. As the labeling was conducted before preprocessing, exclamatory marks and emoticons were also taken into consideration. In addition to manual inspection of these labeled tweets, the tweets were investigated through Bag of Words, top words and their corresponding frequencies for each class, and conjugated words.



**Figure 2.** Distribution of the optimal number of topics with the top 30 salient words. PC1: principal component 1; PC2: principal component 2.



**Textbox 5.** Algorithm 1: polarity assignment to the topic found from the latent Dirichlet allocation Model.

#### Algorithm 1

- Input: topic list
- Output: aspect for each topic
- /\* from nltk.corpus import opinionLexicon
- topicList = Topic List
- sentimentList =
- Topic wise sentiment List \*/
  - positiveWordFromLexicon = set(opinionLexicon.positive()) /\* Positive words from opinionLexicon \*/
  - negativeWordFromLexicon = set(opinionLexicon.negative()) /\* Negative words from opinionLexico \*/
  - positiveWords = getTopWords(labeledPositiveTweets) /\* Positive words from Positive Labeled tweets \*/
  - negativeWords = getTopWords(labeledNegativeTweets) /\* Negative words from Positive Labeled tweets \*/
  - ForEach *topic* in *topicList* do
    - positiveCount = 0
    - negativeCount = 0
    - neutralCount = 0
  - ForEach *word* in *topic.WordList* do
    - if (*word* in positiveWordFromLexicon) OR (*word* in positiveWords) then:
      - positiveCount = 0
    - if (*word* in negativeWordFromLexicon) OR (*word* in negativeWords) then:
      - negativeCount += 1
    - if *word* not in (negativeWordFromLexicon AND negativeWords AND positiveWordFromLexicon AND positiveWords) then:
      - neutralCount += 1
    - if (positiveCount > negativeCount) AND (positiveCount > neutralCount) then:
      - sentimentList[topic:Id] = Positive
    - if (negativeCount > positiveCount) AND (negativeCount > neutralCount) then:
      - sentimentList[topic:Id] = Negative
    - if (neutralCount > positiveCount) AND (neutralCount > negativeCount) then:
      - sentimentList[topic:Id] = Neutral

### Ethics Approval

The research protocol was approved by the School Research Ethics Panel of Allied Health, Faculty of Health, Education, Medicine, and Social Care, Anglia Ruskin University (AH-SREP-19-055).

### Results

The data set was analyzed from the NLP perspective as well as using a qualitative approach.

#### Analysis of Unlabeled Data

The analysis of unlabeled data was 2-fold. First, we analyzed the data by converting them into topics using the Gensim LDA

Model and exploring the aspects using the NLTK opinion dictionary. Second, we used the VADER model to analyze the data set and define the polarity of each tweet.

From [Multimedia Appendix 1](#), we can see the number of topics and their corresponding coherence values, from which we can observe that the coherence values increase with the number of topics, with the highest value of 0.54 at the index numbers 24, 27, and 29. However, along with the higher coherence values, the lower the number of topics and the lower the overlaps among the topics, the better the topic distribution. The topic distribution visualized in [Figure 2](#) shows that, with the number of topics being 28, the LDA Model overlaps lower than the others. Therefore, we selected the optimal number of the topic index to be 27, which implies that the optimal number of topics in

this model index is 28 (the highest coherence is 0.54 with fewer overlaps among the topics) [69]. Moreover, in Figure 2, we show the top 30 salient terms of the data set (on the right) along with the intertopic distance map (on the left).

Figure 3 shows the word cloud of the top words for each topic, along with their broader polarity, where the proportion of positive, negative, and neutral polarity are 28% (8/28), 61% (17/28), and 11% (3/28), respectively. From our analysis, some of the top words from the negative polarity were *awful, bad, lose, lost, tax, pay, money, worker, life, family, social, police, time, meeting, call, care, covid, health, labor, year, week, and time*. Many informal (slang) words were also found in the negative topics. The top words from positive polarity were *business, challenge, productivity, zoom, meeting, find, group, app, check, service, food, worker, love, great, amazing, school, child, day, and story* and, from neutral polarity, the top words were *vote, poll, coming, town, tonight, game, team, news, and*

*today*. Some of the other words from the neutral topics in Figure 3 were *awful, meeting, time, and day*. However, similar terms can be used in different contexts, and the topic distribution provides some idea about the contexts of the topics; for example, topic 5 indicates financial aspects, topic 0 indicates web-based meetings and conferences, and topic 7 is about sports.

In the next step, each tweet was analyzed using VADER. It considered the semantic and contextual meanings of a tweet and calculated a value that showed the intensity of the tweet's sentiment. On the basis of this value, we tagged each tweet as positive, negative, or neutral. In Figure 4, we visualize the number of tweets according to sentiment intensity. Most tweets were within  $-0.30$  to  $+0.30$ , which shows the low intensity of negative or positive sentiments. It also shows that a considerable number of tweets expressed a neutral sentiment (intensity value from  $-0.05$  to  $+0.05$ ).

Figure 3. Top words in each topic within the unlabeled data set using latent Dirichlet allocation Mallet.



**Figure 4.** Sentiment intensity of the unlabeled tweets using Valence Aware Dictionary for Sentiment Reasoner.

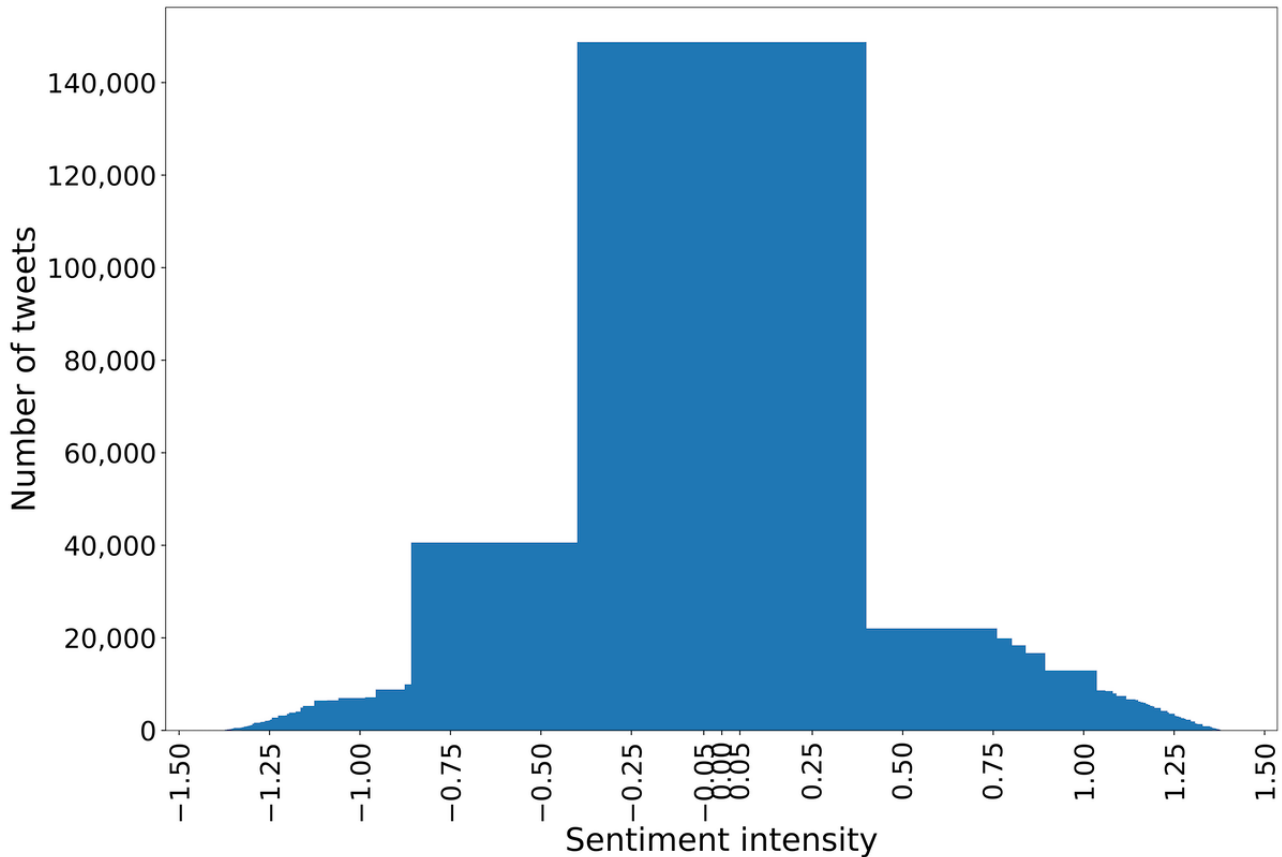


Table 1 shows the intensity level of the tweets, whereas, in Figure 5, we can observe that 48.72% (716,843/1,471,209) of the tweets had positive sentiment, 39.18% (576,462/1,471,209) had negative sentiment, and 12.09% (177,904/1,471,209) indicated neutral sentiments. From Table 1, we observe that most positive tweets have an intensity value between 0.50 and 0.75. The next maximum number of tweets had an intensity between 0.75 and 1.0, which signifies the higher intensity of positive sentiment. By contrast, the maximum number of negative sentiment–showing tweets was within –0.50 to –0.25, which shows the medium level of intensity of negative emotion.

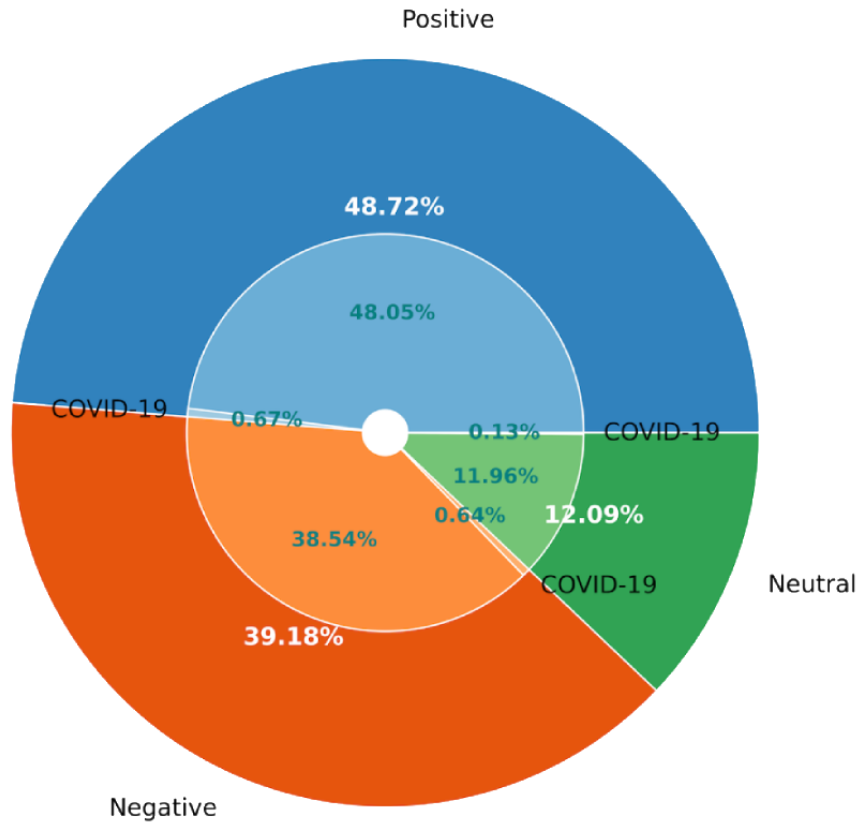
In Figure 5, we also show that 1.44% (21,171/1,471,209) of the tweets mentioned the term “Covid,” among which 0.67% (9854/1,471,209) were positive tweets, 0.64% (9426/1,471,209) were negative tweets, and 0.13% (1891/1,471,209) were neutral tweets.

The negative topics found in the LDA Model were 61% (17/28), and Figure 5 shows that approximately half of the tweets (716,843/1,471,209, 48.72%) expressed positive sentiment. This indicates that, although the negative tweets might be comparatively smaller in number, they imply more diverse subjects than positive tweets relating to work-related mental health based on the analysis conducted using VADER.

**Table 1.** Number of tweets in each quadrant of sentiment intensity.

Group of sentiment intensity	Tweets, n (%)
–1.0 to –0.75	138,659 (9.42)
–0.75 to –0.50	158,974 (10.8)
–0.50 to –0.25	180,014 (12.23)
–0.25 to –0.05	99,028 (6.73)
–0.05 to 0.05	177,691 (12.07)
0.05 to 0.25	105,703 (7.18)
0.25 to 0.50	199,827 (13.58)
0.50 to 0.75	207,439 (14.09)
0.75 to 1.0	203,874 (13.85)

**Figure 5.** The proportion of positive, negative, and neutral unlabeled tweets using Valence Aware Dictionary for Sentiment Reasoner.



**Analysis of Labeled Data**

On the basis of the manually labeled data set, more than half of the tweets (1952/3200, 61%) were categorized or found to be negative; only one-third of the tweets (1088/3200, 34%) expressed the sentiment regarding work-related mental health positively. An exceedingly small portion of the tweets (160/3200, 5%) had an overall neutral sentiment associating work and mental health.

Subsequent to preprocessing and the elimination of the keywords, a digital inspection of this manually labeled data set

revealed the 10 most frequent single words and their frequencies for each polarity, as shown in Tables 2, 3, and 4. These figures also show the words that are strongly connected with work or mental health at work either explicitly or implicitly. Certain words, such as *time, like, life, day, get, need, and much*, appeared with high frequencies in both positive and negative tweets, which is not surprising linguistically. Therefore, it is important to put these words in context, for which these words (Tables 2, 3, and 4) were put together with their previous or next words in Multimedia Appendices 2, 3, and 4.

**Table 2.** Top 10 frequent words from positive tweets with their corresponding frequencies.

Words	Frequency, n (%)
Time	93 (13.1)
Day	88 (12.5)
Get	81 (11.5)
Life	74 (10.5)
Good	68 (9.6)
Much	64 (9)
Take	63 (9)
Like	60 (8.5)
Year	58 (8.2)
Need	57 (5)

**Table 3.** Top 10 frequent words from negative tweets with their corresponding frequencies.

Words	Frequency, n (%)
Time	153 (13.6)
Like	139 (12.3)
Get	138 (12.1)
Life	136 (12)
Day	117 (10.3)
Much	105 (9.3)
People	88 (7.8)
Need	88 (7.8)
Week	85 (7.5)
One	83 (7.3)

**Table 4.** Top 10 frequent words from neutral tweets with their corresponding frequencies.

Words	Frequency, n (%)
Health	10 (13)
People	9 (11.7)
Get	9 (11.7)
Employee	8 (10.4)
Good	8 (10.4)
Make	7 (9.1)
New	7 (9.1)
Life	7 (9.1)
Year	6 (7.8)
Need	6 (7.8)

For example, time was associated with work and mental health, appearing in both positive and negative tweets. Some of the other words conjugated with time in positive tweets were *vacation, season, lucrative, family, financial, burning, spending, kid, life, care, and tough*, whereas, in negative tweets, time was conjugated with words such as *lost, missed, saynotostress (say no to stress), fun, bribed, part time, extra time, hell, and full time*. From this digital inspection, it seems that people talk about the time that is related to their *family, vacation, life, financial situation, or kid* while positively associating work or mental health, and negative emotions were mostly linked to *part time, full time, or extra time*. Low income and insecurity about the source of income over time could be some of the underlying reasons for such negative sentiments.

Some other frequent words such as *year* in positive tweets and *week* and *day* in negative tweets also imply time. Some of the words conjugated with year were *parent, people, two years hard, half, best, and passed*. In negative tweets, some of the conjugated words were *day long, long traffic, Friday, emotional, delay, project, migraine, packed, insane, blood, exhausted, working week, funeral, and unreasonable*. Some slang words were also conjugated in the negative tweets. Analyzing the top words, their adjacent words in terms of meaning, and top words

with the conjugated words, time was the most talked about topic concerning work and mental health in the labeled data.

The second most frequent word was *like*. Although it was comparatively less frequent in positive tweets, it was frequent in negative tweets. In positive tweets, some of the conjugated words were *kindness, laughed like never, issue, loss, feeling, and death*. In contrast, some of the conjugated words in negative tweets were *feel, diagnosis, opportunity, nothing, terrible, isolated, and quite*. The word *feel* was also prevalent in both categories; negative tweets linked feeling with words such as *need, never, fantasy, and fresh*. In positive tweets, it was related to words such as *saw, anything, fit, and floating*.

*Get* was the third most frequent word in both the positive and negative categories. However, in positive tweets, it was related to words such as *relieved, deported, break paralegal, need mode, get done, blessed, get loved, and comfort touch*, whereas, in negative tweets, it was related to words such as *allowing share, get rid, tonight, bad, mean, grade bad, trying sleep, enjoy, money sad, and tension angry*.

*Life* was another frequent word. In positive tweets, it was related to words such as *balance, touch, goal, home, personal, music, and rest*. These imply the issue of work-life balance. Some of the conjugated words in the negative tweets were *personal,*

*private, entire, loss, stressful, release, professional, tough, know life lately, and normal life bam.*

Positive tweets also contained words such as *good, help, and love* with high frequency, whereas negative tweets contained words such as *feel, thing, make, and people*.

Neutral tweets mostly contained words such as *healthy, employee, good, make, need, sleep, and workplace*, and the most common n-grams of them were *belief healthy behaviour, walkout, difficult people, black people drug, red people, attraction get ticket, break paralegal, employee appreciate, minority, impacted, good, idea make workplace, binge, labour life, intelligence life balance, marriage, weekend, enforcement, and break*.

The most frequent words and their n-grams were used to express opposite opinions; this study performed manual qualitative assessments of the labeled tweets. Among the labeled tweets, many claimed that their lifestyle was full of stress—from *dawn to dusk, at home, and at work*. This observation was in agreement with the n-gram analysis. Similar to the top word analysis, issues of burnout were noticeable during the manual assessment.

For some people, the holidays increase their stress because of the pressure of workload after the holidays—such rationale for this sentiment could not be perceived through the analytical tools.

Commuting from home to the office was found to be stressful for many tweeters, which is also in agreement with the n-gram analysis.

Tweeters often directly communicate their physiological and mental symptoms linked to work, such as *back pain, headache, nervousness, anxiety, sleep deprivation, and loss of appetite*, in words. Tweeters attempt to reduce work stress through smoking and substance use, which was evident in the n-gram-based analysis as well.

There were also many tweets only stating work stress using upper case, emphasizing intense stress. The qualitative inspection of this data set revealed work stress in the sporting world, education, and health professions as well as concerning elections, none of which was highlighted through n-gram-based analysis.

Work-home balance and marital and relationship issues were spotted in the labeled tweets, as highlighted by the n-gram analysis as well. High-risk jobs such as firefighting were linked to work stress and marital life. Remarks on work stress and less time spent with spouses, leading to an irritable state of mind, were also noticeable. Tweeters (ie, users of Twitter) also spoke about bad moods as a result of work stress in general. Tweeters expressed confusion regarding the cause and effect of work stress and exhaustion.

Tweeters spoke of suicide, not as direct thoughts or passive suicidal ideation but rather as suicidal incidents associated with stress within their social group. It was not clear whether the person in question belonged to their own work group or professional fraternity or was someone they knew whose suicide was linked to work.

The manual qualitative analysis also revealed some interesting coping mechanisms; for example, being inappropriate in tweet comments was mentioned as one of the coping strategies for work stress. Positive tweets on work-related mental health also mostly concentrated on coping strategies such as listening to music by certain trending artists, gaming, watching a situation comedy on television, traveling, spending time with friends, and social drinking. Tweeters also shared auto-suggestions such as letting go of work stress to prioritize life and a philosophy on impermanent life. The practice of mindfulness and relaxation was encouraged by the positive tweeters to relieve work stress.

Different studies and research findings contributed to most of the neutral tweets in the labeled data set. Owing to the careful selection of the keywords, ambiguity regarding the broader polarity of sentiments was restricted. However, some of the neutral-labeled tweets contained recommendations on how to deal with work stress as they could not be clearly categorized as positive.

Analyzing mood- and mental health-related issues is not a simple task as emotions are often mixed or ambivalent. It is possible to describe mixed emotions as the simultaneous experience of different combinations of opposing emotions, and positive and negative emotions can occur simultaneously, as is evident in this section as well.

## Discussion

### Challenges

In 2019 to 2020, the economic expense of lost working days was estimated to be >GBP 16 billion (US \$19.3 billion) [1]. The official numbers suggest an expenditure of GBP 3.5 billion (US \$4.2 billion) borne by the UK government (at the end of the day, the UK population) because of work-related injuries and ill health [1]. Ignoring such prevalence of work-related poor health adds more to the national disease burden, which may not be reflected as a direct cost occurring from the unhealthy and toxic working conditions. In 2018 to 2019, employers faced a loss of GBP 3.2 billion (US \$3.9 billion) as workers were off sick for reasons attributable to work itself [1]. The pandemic and the socioeconomic consequences in the postpandemic period have created a more unsettling environment. An unhappy workforce hinders the overall growth and productivity of any organization [70,71]. However, most of the economic burden of work-related mental and physical ill health is borne by the worker and their friends and family, who also bear the emotional hardship of such adversity. Therefore, breaking the cycle of poor health outcomes is a point of interest at a national, organizational, and personal level.

Nowadays, mental health and well-being have become new buzzwords, which can be thought of as both positive and negative signs. Recognizing the problem is always the first step. More visibility and open discussions around mental health issues may open the door to reducing stigma and breaking the taboo on this subject. Self-awareness and strategic approaches from the employer may significantly reduce the mental health effects arising because of the workplace. Many commercial and nonprofit organizations have been trying to pave their way into

this market by delivering business-to-business support, such as Minddistrict [72], the Health Assured employee assistance program [73], and Nuffield Health [74].

AI-enabled, data-driven approaches have already shown promising results in predicting quitters in the workforce [75]. Zegami [76], launched at the University of Oxford, has been working on a tool to identify unhappy employees using a wide range of data such as age, salary, benefits, and work location [77].

The sentiment around the workplace may be expressed in words, body language, physiological parameters, acts and actions, or responses to others. However, it can be challenging to prove the root cause of the issue when the background narrative is also part of the problem—psychological, socioeconomic, cultural, and person-specific profiles may define how such sentiment is perceived and expressed. Moreover, one-third of work-related ill health is due to musculoskeletal disorders [1]. There is a consensus among clinicians and occupational health experts that musculoskeletal disorders are linked to mental health issues, the mechanism of which is yet to be unraveled [78,79]. Work-related ill health, linked to psychosomatic effects, cultivates a vicious cycle of habits, such as work-related anxiety and stress causing insomnia, which causes headache and increased intake of stimulants, which causes more sleepless nights, which causes stress, and the cycle continues.

### Framework for Combining Existing Technologies

The complex and dynamic relationships (Figure 6) among work patterns, lifestyle, individualized work patterns, occupational hazards, health, and well-being are all intertwined with the workplace sentiments. Harnessing the power of AI and big data analytics, a descriptive model can be constructed to unravel the

shared responsibility among stakeholders (eg, the individual, employer, and industry).

This study sketches a framework to understand the vulnerable zone of work-related sentiments to facilitate better work practice and health outcomes (Figure 7). In Figure 7, the global descriptors within the problem space can be broadly categorized into three domains as follows: (1) industry-specific elements arising from the nature of the occupation (eg, clinicians at night shifts); (2) organization-specific workplace culture (eg, social drinking culture); and (3) societal descriptors such as the Gross Happiness Index, political stability, crime rate, employment rate, living expenses, economy, social inclusion, diversity, and tolerance, the data for which can be aggregated from national and global databases. These global descriptors directly influence the local descriptors forming an unbreakable pattern; however, the impact can vary widely from person to person.

Developing an understanding of how individuals feel about such external variables may not be reflected in numbers (see the Introduction section) or explicitly expressed in words (see this section). Deeper insights into the factors responsible for work-related sentiments through the local descriptor need to be coproduced by engaging the workforce in the process. Cross-referencing the local descriptors of the descriptive model with real-time lifestyle tracking along with global descriptors of mental health disorders (data to be collected through electronic health records) can inform AI-enabled predictive models. Evidence-based prediction will empower stakeholders to make informed decisions, design appropriate preventive measures to reduce the prevalence of work-related ill health, and introduce effective interventions to keep the workforce healthy and happy.

**Figure 6.** Dynamic relationship between work life and health.

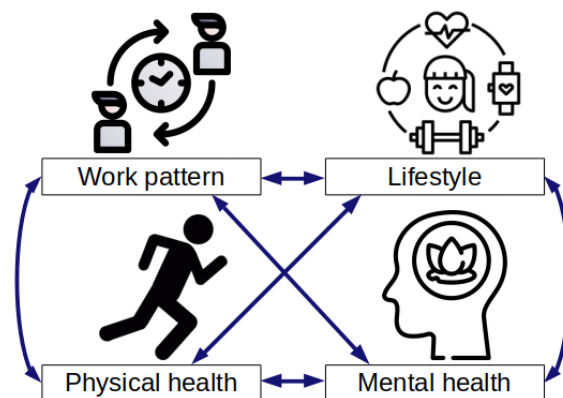
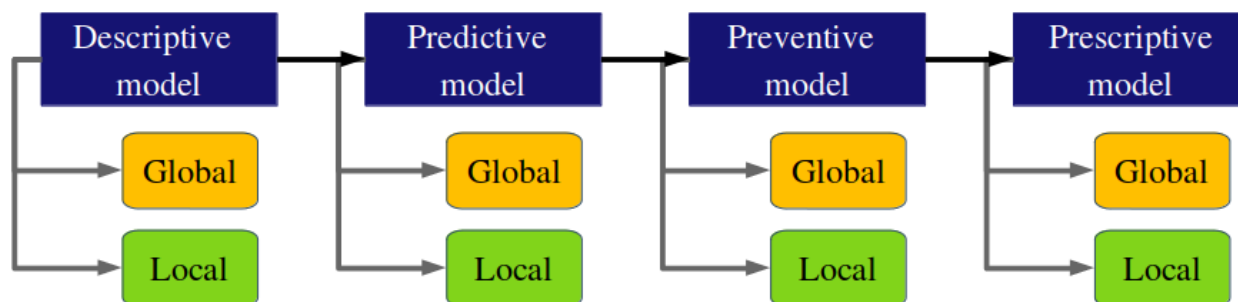




Figure 7. High-level architecture of artificial intelligence-enabled mental health support for the workforce.



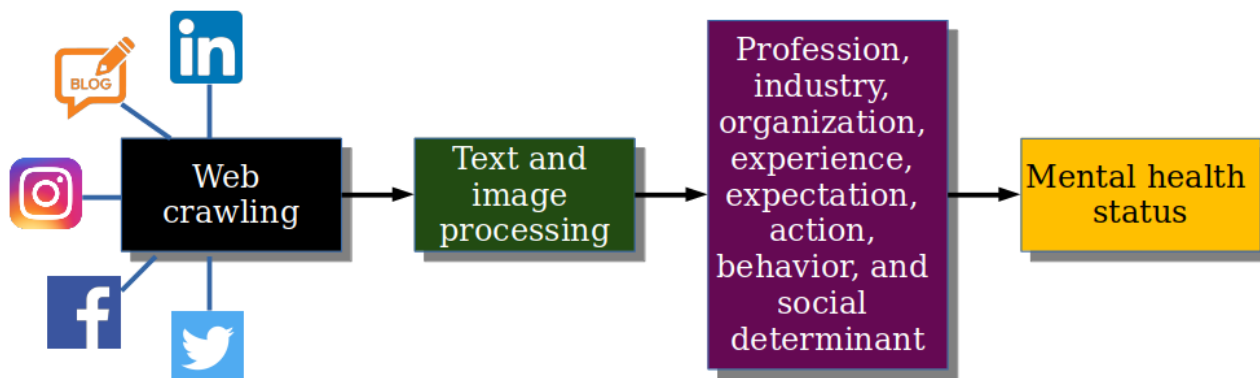
### The Discourse Regarding Ultrasmart Sentiment Detection

The habit and necessity to share personal information on public platforms and the control over the outcome of sharing such information are 2 sides of the same coin in the era of datafication. The consequence of carefree users relinquishing rights to privacy increases the risk of invasion of privacy of users who are not even engaged with the process. The end users are often unaware of digital traps such as “Cookies” and “personalization” and the emerging and inconceivable powers of AI with invisible data links, resulting in an accidental discovery of an untold (fabricated) story and data breaches. Such an outcome often results from the negligence of the data controller and data processor. Examples are evident in the literature [80-82].

This study presents a hypothetical data map of how publicly available disintegrated data can reveal the “digital” mental health

of an employee (Figure 8). An employer can analyze the time-stamped digital footprint to evaluate and predict the performance of employees; however, it is questionable without full consent on data collection and processing and without strong data governance and data ethics in place. Such profiling, as shown in Figure 8, can be constructed in a third-party organization or even via bots. Although the General Data Protection Regulation protects such personal data mining and processing up to a certain degree within the United Kingdom and European Union region, care should be taken to safeguard the vulnerable workforce, which is already experiencing poor mental health conditions. Although this study advocates for a more comprehensive approach using a wide range of technological tools to better perceive the occupational health concerns and improve the health and well-being of the workforce, care should be taken while using these emerging techniques to avoid the creation of a Pandora’s box of digital surveillance.

Figure 8. Self-imposed digital surveillance without a security breach.



### Research Limitations and Future Work

Twitter is one of the most popular microblogs for researchers, where millions of people express their feelings. Owing to its character limit, users are bound to write specific data, which helps in analyzing them. Moreover, Twitter provides application programming interfaces for researchers to fetch and analyze archived tweets in an enormous number [83]. However, not all working adults are on social media, not all tweeters are verbally expressive, and a single event of microblogging can be insufficient and inconclusive to perceive deeper emotions. Moreover, millions of users prefer other microblogging sites to express their emotions. Therefore, despite having many advantages for researchers, using Twitter alone as a data source is a limitation. Although this study considered public tweets,

the research was subject to approval from one of the internal ethics committees. They provided approval on the grounds that tweets could not be directly quoted on the paper, and no tweeter would be tracked. Therefore, we could not include and analyze any personal data such as usernames and locations. A more detailed context for the inclusion of such data can be found elsewhere [84-86]. However, this study used an extensive data fetching process to avoid such limitations. Furthermore, considering the ethical limitations, a case study will require us to recruit participants, which needs to be carefully done as a constant awareness of being analyzed may hamper the spontaneity of the tweets. The inclusion of geographical analysis and longitudinal observation of the same tweeter in the future can provide richer insights and a better context of work-related sentiments to unveil a pattern, if it exists. However, data

governance in data processing should be prioritized, and constant caution should be exercised so that users' privacy is not hampered. Although including a case study and geographical analysis is beyond the scope of this study, we are encouraged to do so in the future.

## Conclusions

This paper presented a synopsis of the current state of work-related mental health, its ripple effect beyond working hours, and how it is being dealt with. The examination of >1 million tweets in the aforementioned sections shows distinct aspects of work-related sentiments that people are willing to

share on social media. The key findings of this study are outlined in [Textbox 6](#).

Analyzing thousands of tweets, this study exposed the aspects that the workforce is concerned about, which are in agreement with the LFS survey to some extent. However, the subject matter requires a more integrated approach. Moreover, irony, sarcasm, emojis, multipolarity, and word ambiguity are some of the challenging aspects of the NLP domain that require further enhancement. Therefore, we emphasize exploring how technology-driven solutions can support the unraveling of the dynamically intertwined relationship among work, mental and physical health, and lifestyle to protect our workforce and transform the workplace.

**Textbox 6.** Key findings of this study.

### Key findings

- The type of words people use to express work-related sentiments were identified.
- Colloquialism and eloquence were both observed in Twitter expressions regarding work, whether expressing grief or happiness.
- An overlapping set of words in both positive and negative tweets signifies upbeat and, by contrast, undesirable feelings affecting the same entities (day, life, and time) in their lives.
- Time variables (*day*, *time*, and *week*), *life*, and *need* were the predominant words in both positive and negative tweets—as is evident from the trigrams. The topic analysis provided a sense of contextual narrative across sentiments. For example, topic 5 indicates financial aspects, topic 0 indicates internet-based meetings and conferences, and topic 7 is about sports.

## Acknowledgments

This work was supported in part by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), and in part by an InnoHK Project at the Hong Kong Centre for Cerebro-cardiovascular Health Engineering (COCHE). DAC is an Investigator in the Pandemic Sciences Institute, University of Oxford, Oxford, UK. MHT is supported by an Engineering and Physical Sciences Research Council Healthcare Technologies Challenge Award (EP/N020774/1). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health, InnoHK – ITC, Engineering and Physical Sciences Research Council, or the University of Oxford.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Coherence values of the latent Dirichlet allocation Model for a number of topics from 2 to 31.

[\[PDF File \(Adobe PDF File\), 105 KB - formative\\_v6i9e30113\\_app1.pdf\]](#)

### Multimedia Appendix 2

Top words from positive tweets with previous and next words (trigrams).

[\[PDF File \(Adobe PDF File\), 433 KB - formative\\_v6i9e30113\\_app2.pdf\]](#)

### Multimedia Appendix 3

Top words from negative tweets with previous and next words (trigrams).

[\[PDF File \(Adobe PDF File\), 452 KB - formative\\_v6i9e30113\\_app3.pdf\]](#)

### Multimedia Appendix 4

Top words from neutral tweets with previous and next words (trigrams).

[\[PDF File \(Adobe PDF File\), 320 KB - formative\\_v6i9e30113\\_app4.pdf\]](#)

## References

1. Health and safety executive homepage. HSE. 2022. URL: <https://www.hse.gov.uk> [accessed 2022-06-08]

2. Labour Force Survey. Office for National Statistics. 2020. URL: <https://www.ons.gov.uk/surveys> [accessed 2022-06-08]
3. Leach C, Finning K, Kerai G, Vizard T. Coronavirus and depression in adults, Great Britain: July to August 2021. Office for National Statistics. 2021 Oct 1. URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/coronavirusanddepressioninadultsgreatbritain/julytoaugust2021> [accessed 2022-06-08]
4. The five-year forward view for mental health. NHS England. 2016 Feb. URL: <https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf> [accessed 2022-06-08]
5. Mental health and substance use. World Health Organization. URL: <https://www.who.int/teams/mental-health-and-substance-use/overview> [accessed 2022-06-08]
6. Hampson E, Soneji U, Jacob A, Mecu B, Gahan HM. Mental health and employers: the case for investment. Monitor Deloitte. 2017 Oct. URL: <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/public-sector/deloitte-uk-mental-health-employers-monitor-deloitte-oct-2017.pdf> [accessed 2022-08-19]
7. Stevenson LD, Farmer P. Thriving at Work: a review of mental health and employers. GOV.UK. 2017 Oct 26. URL: <https://www.gov.uk/government/publications/thriving-at-work-a-review-of-mental-health-and-employers> [accessed 2022-08-19]
8. Kalmady SV, Greiner R, Agrawal R, Shivakumar V, Narayanaswamy JC, Brown MR, et al. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. NPJ Schizophr 2019 Jan 18;5(1):2 [FREE Full text] [doi: [10.1038/s41537-018-0070-8](https://doi.org/10.1038/s41537-018-0070-8)] [Medline: [30659193](https://pubmed.ncbi.nlm.nih.gov/30659193/)]
9. Chivilgina O, Wangmo T, Elger BS, Heinrich T, Jotterand F. mHealth for schizophrenia spectrum disorders management: a systematic review. Int J Soc Psychiatry 2020 Nov;66(7):642-665. [doi: [10.1177/0020764020933287](https://doi.org/10.1177/0020764020933287)] [Medline: [32571123](https://pubmed.ncbi.nlm.nih.gov/32571123/)]
10. Günther MP, Kirchebner J, Lau S. Identifying direct coercion in a high risk subgroup of offender patients with schizophrenia machine learning algorithms. Front Psychiatry 2020;11:415 [FREE Full text] [doi: [10.3389/fpsy.2020.00415](https://doi.org/10.3389/fpsy.2020.00415)] [Medline: [32477188](https://pubmed.ncbi.nlm.nih.gov/32477188/)]
11. Cai X, Xie D, Madsen K, Wang Y, Bögemann SA, Cheung E, et al. Generalizability of machine learning for classification of schizophrenia based on resting-state functional MRI data. Hum Brain Mapp 2020 Jan;41(1):172-184 [FREE Full text] [doi: [10.1002/hbm.24797](https://doi.org/10.1002/hbm.24797)] [Medline: [31571320](https://pubmed.ncbi.nlm.nih.gov/31571320/)]
12. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: [10.2196/11334](https://doi.org/10.2196/11334)] [Medline: [30681968](https://pubmed.ncbi.nlm.nih.gov/30681968/)]
13. Zhang J, Richardson JD, Dunkley BT. Classifying post-traumatic stress disorder using the magnetoencephalographic connectome and machine learning. Sci Rep 2020 Apr 03;10(1):5937 [FREE Full text] [doi: [10.1038/s41598-020-62713-5](https://doi.org/10.1038/s41598-020-62713-5)] [Medline: [32246035](https://pubmed.ncbi.nlm.nih.gov/32246035/)]
14. Păsăreanu CR, Andersson G, Dobrea A. Attention-deficit/hyperactivity disorder mobile apps: a systematic review. Int J Med Inform 2020 Jun;138:104133 [FREE Full text] [doi: [10.1016/j.jimedinf.2020.104133](https://doi.org/10.1016/j.jimedinf.2020.104133)] [Medline: [32283479](https://pubmed.ncbi.nlm.nih.gov/32283479/)]
15. Harkko J, Sumanen H, Pietiläinen O, Piha K, Mänty M, Lallukka T, et al. Socioeconomic differences in occupational health service utilization and sickness absence due to mental disorders: a register-based retrospective cohort study. Int J Environ Res Public Health 2020 Mar 20;17(6):2020-2021 [FREE Full text] [doi: [10.3390/ijerph17062064](https://doi.org/10.3390/ijerph17062064)] [Medline: [32244960](https://pubmed.ncbi.nlm.nih.gov/32244960/)]
16. Okuda Y, Iwasaki S, Deguchi Y, Nitta T, Mitake T, Sakaguchi A, et al. Burnout and occupational stressors among non-medical occupational health staff. Occup Med (Lond) 2020 Mar 12;70(1):45-51 [FREE Full text] [doi: [10.1093/occmed/kqz160](https://doi.org/10.1093/occmed/kqz160)] [Medline: [31860099](https://pubmed.ncbi.nlm.nih.gov/31860099/)]
17. Milner A, Maheen H, Currier D, LaMontagne A. Male suicide among construction workers in Australia: a qualitative analysis of the major stressors precipitating death. BMC Public Health 2017 Jun 19;17(1):584-586 [FREE Full text] [doi: [10.1186/s12889-017-4500-8](https://doi.org/10.1186/s12889-017-4500-8)] [Medline: [28629352](https://pubmed.ncbi.nlm.nih.gov/28629352/)]
18. Gradus J, Rosellini A, Horváth-Puhó E, Street A, Galatzer-Levy I, Jiang T, et al. Prediction of sex-specific suicide risk using machine learning and single-payer health care registry data from Denmark. JAMA Psychiatry 2020 Jan 01;77(1):25-34 [FREE Full text] [doi: [10.1001/jamapsychiatry.2019.2905](https://doi.org/10.1001/jamapsychiatry.2019.2905)] [Medline: [31642880](https://pubmed.ncbi.nlm.nih.gov/31642880/)]
19. Lin G, Nagamine M, Yang S, Tai Y, Lin C, Sato H. Machine learning based suicide ideation prediction for military personnel. IEEE J Biomed Health Inform 2020 Jul;24(7):1907-1916. [doi: [10.1109/jbhi.2020.2988393](https://doi.org/10.1109/jbhi.2020.2988393)]
20. Hassanpour S, Tomita N, DeLise T, Crosier B, Marsch LA. Identifying substance use risk based on deep neural networks and Instagram social media data. Neuropsychopharmacology 2019 Feb 24;44(3):487-494 [FREE Full text] [doi: [10.1038/s41386-018-0247-x](https://doi.org/10.1038/s41386-018-0247-x)] [Medline: [30356094](https://pubmed.ncbi.nlm.nih.gov/30356094/)]
21. Geraci J, Wilansky P, de Luca V, Roy A, Kennedy JL, Strauss J. Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression. Evid Based Ment Health 2017 Aug 24;20(3):83-87 [FREE Full text] [doi: [10.1136/eb-2017-102688](https://doi.org/10.1136/eb-2017-102688)] [Medline: [28739578](https://pubmed.ncbi.nlm.nih.gov/28739578/)]
22. Powell AC, Torous JB, Firth J, Kaufman KR. Generating value with mental health apps. BJPsych Open 2020 Feb 05;6(2):e16 [FREE Full text] [doi: [10.1192/bjo.2019.98](https://doi.org/10.1192/bjo.2019.98)] [Medline: [32019619](https://pubmed.ncbi.nlm.nih.gov/32019619/)]
23. Arenas-Castañeda PE, Aroca Bisquert F, Martínez-Nicolas I, Castillo Espíndola LA, Barahona I, Maya-Hernández C, et al. Universal mental health screening with a focus on suicidal behaviour using smartphones in a Mexican rural community: protocol for the SMART-SCREEN population-based survey. BMJ Open 2020 Jul 19;10(7):e035041 [FREE Full text] [doi: [10.1136/bmjopen-2019-035041](https://doi.org/10.1136/bmjopen-2019-035041)] [Medline: [32690505](https://pubmed.ncbi.nlm.nih.gov/32690505/)]

24. Görlich D, Faldum A. Implementing online interventions in ICare: a biostatistical perspective. *Internet Interv* 2019 Apr;16:5-11 [FREE Full text] [doi: [10.1016/j.invent.2018.12.004](https://doi.org/10.1016/j.invent.2018.12.004)] [Medline: [30775259](https://pubmed.ncbi.nlm.nih.gov/30775259/)]
25. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial intelligence-assisted online social therapy for youth mental health. *Front Psychol* 2017;8:796 [FREE Full text] [doi: [10.3389/fpsyg.2017.00796](https://doi.org/10.3389/fpsyg.2017.00796)] [Medline: [28626431](https://pubmed.ncbi.nlm.nih.gov/28626431/)]
26. Drissi N, Ouhbi S, Janati Idrissi MA, Ghogho M. An analysis on self-management and treatment-related functionality and characteristics of highly rated anxiety apps. *Int J Med Inform* 2020 Sep;141:104243 [FREE Full text] [doi: [10.1016/j.ijmedinf.2020.104243](https://doi.org/10.1016/j.ijmedinf.2020.104243)] [Medline: [32768994](https://pubmed.ncbi.nlm.nih.gov/32768994/)]
27. Bakker D, Kazantzis N, Rickwood D, Rickard N. Development and pilot evaluation of smartphone-delivered cognitive behavior therapy strategies for mood- and anxiety-related problems: MoodMission. *Cognit Behav Pract* 2018 Nov;25(4):496-514 [FREE Full text] [doi: [10.1016/j.cbpra.2018.07.002](https://doi.org/10.1016/j.cbpra.2018.07.002)]
28. Boonstra T, Werner-Seidler A, O'Dea B, Larsen M, Christensen H. Smartphone app to investigate the relationship between social connectivity and mental health. In: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2017 Presented at: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 11-15, 2017; Jeju, Korea (South). [doi: [10.1109/embc.2017.8036818](https://doi.org/10.1109/embc.2017.8036818)]
29. Callan JA, Dunbar Jacob J, Siegle GJ, Dey A, Thase ME, DeVito Dabbs A, et al. CBT MobileWork©: user-centered development and testing of a mobile mental health application for depression. *Cogn Ther Res* 2020 Sep 25;45(2):287-302. [doi: [10.1007/s10608-020-10159-4](https://doi.org/10.1007/s10608-020-10159-4)]
30. Kannisto K, Korhonen J, Adams C, Koivunen M, Vahlberg T, Välimäki MA. Factors associated with dropout during recruitment and follow-up periods of a mHealth-based randomized controlled trial for mobile.net to encourage treatment adherence for people with serious mental health problems. *J Med Internet Res* 2017 Feb 21;19(2):e46 [FREE Full text] [doi: [10.2196/jmir.6417](https://doi.org/10.2196/jmir.6417)] [Medline: [28223262](https://pubmed.ncbi.nlm.nih.gov/28223262/)]
31. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv* 2017 Dec;10:39-46 [FREE Full text] [doi: [10.1016/j.invent.2017.10.002](https://doi.org/10.1016/j.invent.2017.10.002)] [Medline: [30135751](https://pubmed.ncbi.nlm.nih.gov/30135751/)]
32. Tania M, Andreev I, Temdee P. Arts and science of digital mental health support. In: Proceedings of the 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON). 2020 Presented at: 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON); Mar 11-14, 2020; Pattaya, Thailand. [doi: [10.1109/ectidamtcon48261.2020.9090741](https://doi.org/10.1109/ectidamtcon48261.2020.9090741)]
33. Revolutionising healthcare by empowering doctors with artificial intelligence. Babylon Health. URL: <https://www.babylonhealth.com/en-gb/ai> [accessed 2021-07-04]
34. Quartet Health homepage. Quartet Health. URL: <https://www.quartethealth.com/> [accessed 2021-07-04]
35. A smarter approach to emotional health. Lyra Health. URL: <https://www.lyrahealth.com/> [accessed 2021-07-04]
36. On-demand mental healthcare. Ginger. URL: <https://www.ginger.com/> [accessed 2021-07-04]
37. Relational agent for mental health. Woebot Health. URL: <https://woebothealth.com/> [accessed 2021-07-04]
38. Longer, fuller lives with digital-first care and research. Huma. URL: <https://huma.com/> [accessed 2021-07-04]
39. This is healthcare without the system. K Health. URL: <https://khealth.com/> [accessed 2021-07-04]
40. Mental healthcare that's precisely right for you. Spring Health. URL: <https://www.springhealth.com/> [accessed 2021-07-04]
41. Shane-Simpson C, Manago A, Gaggi N, Gillespie-Lynch K. Why do college students prefer Facebook, Twitter, or Instagram? Site affordances, tensions between privacy and self-expression, and implications for social capital. *Comput Human Behav* 2018 Sep;86:276-288 [FREE Full text] [doi: [10.1016/j.chb.2018.04.041](https://doi.org/10.1016/j.chb.2018.04.041)]
42. Wang J, Fan Y, Palacios J, Chai Y, Guetta-Jeanrenaud N, Obradovich N, et al. Global evidence of expressed sentiment alterations during the COVID-19 pandemic. *Nat Hum Behav* 2022 Mar 17;6(3):349-358. [doi: [10.1038/s41562-022-01312-y](https://doi.org/10.1038/s41562-022-01312-y)] [Medline: [35301467](https://pubmed.ncbi.nlm.nih.gov/35301467/)]
43. Tsao S, MacLean A, Chen H, Li L, Yang Y, Butt ZA. Public attitudes during the second lockdown: sentiment and topic analyses using tweets from ontario, canada. *Int J Public Health* 2022 Feb 21;67:1604658 [FREE Full text] [doi: [10.3389/ijph.2022.1604658](https://doi.org/10.3389/ijph.2022.1604658)] [Medline: [35264920](https://pubmed.ncbi.nlm.nih.gov/35264920/)]
44. Cuenca-Zaldívar JN, Torrente-Regidor M, Martín-Losada L, Fernández-De-Las-Peñas C, Florencio LL, Sousa PA, et al. Exploring sentiment and care management of hospitalized patients during the first wave of the COVID-19 pandemic using electronic nursing health records: descriptive study. *JMIR Med Inform* 2022 May 12;10(5):e38308 [FREE Full text] [doi: [10.2196/38308](https://doi.org/10.2196/38308)] [Medline: [35486902](https://pubmed.ncbi.nlm.nih.gov/35486902/)]
45. Yin H, Song X, Yang S, Li J. Sentiment analysis and topic modeling for COVID-19 vaccine discussions. *World Wide Web* 2022 Feb 25;25(3):1067-1083 [FREE Full text] [doi: [10.1007/s11280-022-01029-y](https://doi.org/10.1007/s11280-022-01029-y)] [Medline: [35250362](https://pubmed.ncbi.nlm.nih.gov/35250362/)]
46. Saura JR, Ribeiro-Soriano D, Zegarra Saldaña P. Exploring the challenges of remote work on Twitter users' sentiments: from digital technology development to a post-pandemic era. *J Business Res* 2022 Mar;142:242-254. [doi: [10.1016/j.jbusres.2021.12.052](https://doi.org/10.1016/j.jbusres.2021.12.052)]

47. Gibbons J, Malouf R, Spitzberg B, Martinez L, Appleyard B, Thompson C, et al. Twitter-based measures of neighborhood sentiment as predictors of residential population health. *PLoS One* 2019;14(7):e0219550 [FREE Full text] [doi: [10.1371/journal.pone.0219550](https://doi.org/10.1371/journal.pone.0219550)] [Medline: [31295294](https://pubmed.ncbi.nlm.nih.gov/31295294/)]
48. Bryan C, Butner J, Sinclair S, Bryan A, Hesse C, Rose A. Predictors of emerging suicide death among military personnel on social media networks. *Suicide Life Threat Behav* 2018 Aug;48(4):413-430 [FREE Full text] [doi: [10.1111/sltb.12370](https://doi.org/10.1111/sltb.12370)] [Medline: [28752655](https://pubmed.ncbi.nlm.nih.gov/28752655/)]
49. Windsor-Shellard B. Suicide by occupation, England: 2011 to 2015. Office for National Statistics. 2017 Mar 17. URL: <https://bit.ly/3oNMgs6> [accessed 2022-06-08]
50. Sherratt F, Welfare K, Hallowell M, Hoque Tania M. Legalized recreational marijuana: safety, ethical, and legal perceptions of the workforce. *J Constr Eng Manag* 2018 Jun;144(6):04018048 [FREE Full text] [doi: [10.1061/\(asce\)co.1943-7862.0001502](https://doi.org/10.1061/(asce)co.1943-7862.0001502)]
51. Daniulaityte R, Zatreh M, Lamy F, Nahhas R, Martins S, Sheth A, et al. A Twitter-based survey on marijuana concentrate use. *Drug Alcohol Depend* 2018 Jun 01;187:155-159 [FREE Full text] [doi: [10.1016/j.drugalcdep.2018.02.033](https://doi.org/10.1016/j.drugalcdep.2018.02.033)] [Medline: [29669296](https://pubmed.ncbi.nlm.nih.gov/29669296/)]
52. Sung Y, Lee J, Kim E, Choi S. Why we post selfies: understanding motivations for posting pictures of oneself. *Personal Indiv Diff* 2016 Jul;97:260-265 [FREE Full text] [doi: [10.1016/j.paid.2016.03.032](https://doi.org/10.1016/j.paid.2016.03.032)]
53. Kelly Y, Zilanawala A, Booker C, Sacker A. Social media use and adolescent mental health: findings from the UK millennium cohort study. *EClinicalMedicine* 2018 Dec;6:59-68 [FREE Full text] [doi: [10.1016/j.eclinm.2018.12.005](https://doi.org/10.1016/j.eclinm.2018.12.005)] [Medline: [31193561](https://pubmed.ncbi.nlm.nih.gov/31193561/)]
54. Primack B, Shensa A, Sidani J, Whaite E, Lin LY, Rosen D, et al. Social media use and perceived social isolation among young adults in the U.S. *Am J Prev Med* 2017 Jul;53(1):1-8 [FREE Full text] [doi: [10.1016/j.amepre.2017.01.010](https://doi.org/10.1016/j.amepre.2017.01.010)] [Medline: [28279545](https://pubmed.ncbi.nlm.nih.gov/28279545/)]
55. Dhir A, Yossatorn Y, Kaur P, Chen S. Online social media fatigue and psychological wellbeing—a study of compulsive use, fear of missing out, fatigue, anxiety and depression. *Int J Inform Manag* 2018 Jun;40:141-152. [doi: [10.1016/j.ijinfomgt.2018.01.012](https://doi.org/10.1016/j.ijinfomgt.2018.01.012)]
56. Beyens I, Pouwels JL, van Driel II, Keijsers L, Valkenburg PM. The effect of social media on well-being differs from adolescent to adolescent. *Sci Rep* 2020 Jul 01;10(1):10763 [FREE Full text] [doi: [10.1038/s41598-020-67727-7](https://doi.org/10.1038/s41598-020-67727-7)] [Medline: [32612108](https://pubmed.ncbi.nlm.nih.gov/32612108/)]
57. Bell B, Cassarly J, Dunbar L. Selfie-objectification: self-objectification and positive feedback ("Likes") are associated with frequency of posting sexually objectifying self-images on social media. *Body Image* 2018 Sep;26:83-89 [FREE Full text] [doi: [10.1016/j.bodyim.2018.06.005](https://doi.org/10.1016/j.bodyim.2018.06.005)] [Medline: [30005293](https://pubmed.ncbi.nlm.nih.gov/30005293/)]
58. Weinstein E. The social media see-saw: positive and negative influences on adolescents' affective well-being. *New Media Soc* 2018 Feb 21;20(10):3597-3623 [FREE Full text] [doi: [10.1177/1461444818755634](https://doi.org/10.1177/1461444818755634)]
59. Oshio A, Taku K, Hirano M, Saeed G. Resilience and Big Five personality traits: a meta-analysis. *Personal Indiv Diff* 2018 Jun;127:54-60 [FREE Full text] [doi: [10.1016/j.paid.2018.01.048](https://doi.org/10.1016/j.paid.2018.01.048)]
60. Azucar D, Marengo D, Settanni M. Predicting the Big 5 personality traits from digital footprints on social media: a meta-analysis. *Personal Indiv Diff* 2018 Apr;124:150-159 [FREE Full text] [doi: [10.1016/j.paid.2017.12.018](https://doi.org/10.1016/j.paid.2017.12.018)]
61. GitHub Universe: a global developer event. GitHub. URL: <https://github.com/> [accessed 2022-08-19]
62. Edo-Osagie O, Iglesia BD, Lake I, Edeghere O. An evolutionary approach to automatic keyword selection for twitter data analysis. In: *Hybrid Artificial Intelligent Systems. HAIS 2020*. Cham: Springer; 2020.
63. Search Tweets. Twitter. URL: <https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query> [accessed 2022-08-19]
64. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003;3:993-1022.
65. Gensim: topic modelling for humans. Gensim. URL: <https://radimrehurek.com/gensim/> [accessed 2020-09-28]
66. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. *Proc Int AAAI Conf Web Social Media* 2014;8(1):216-225.
67. Kapadia S. Evaluate topic models: Latent Dirichlet Allocation (LDA). *Towards Data Science*. 2019 Aug 19. URL: <https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0> [accessed 2022-05-21]
68. Kurt S. Topic modeling — LDA mallet implementation in Python — Part 2. *Medium*. 2020 Jun 30. URL: <https://medium.com/swlh/topic-modeling-lda-mallet-implementation-in-python-part-2-602ffb38d396> [accessed 2022-05-21]
69. Prabhakaran S. Topic Modeling with Gensim (Python). *Machine Learning Plus*. 2018 Mar 26. URL: <https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/> [accessed 2022-05-21]
70. Johnson S, Robertson I, Cooper CL. *Well-Being Productivity and Happiness at Work*. Cham: Springer International Publishing; 2017.
71. Bellet C, De Neve J, Ward G. Does employee happiness have an impact on productivity? *SSRN J* 2019 Oct 17. [doi: [10.2139/ssrn.3470734](https://doi.org/10.2139/ssrn.3470734)]
72. Digital mental health platform for healthcare providers. *Minddistrict*. URL: <https://www.minddistrict.com/> [accessed 2021-07-04]
73. Health Assured. URL: <https://www.healthassured.org/> [accessed 2021-07-04]

74. Corporate healthcare. Nuffield Health. URL: <https://www.nuffieldhealth.com/corporate-healthcare> [accessed 2021-07-04]
75. Holtom B, Allen D. Better ways to predict who's going to quit. Harvard Business Review. 2019 Aug 16. URL: <https://hbr.org/2019/08/better-ways-to-predict-whos-going-to-quit> [accessed 2022-08-19]
76. End-to-end imaging AI specialists combining unique tools and machine learning services. Zegami. URL: <https://zegami.com/> [accessed 2021-07-04]
77. Boland H. The AI that knows you're about to resign before you do - and warns your boss. The Telegraph. 2019 Oct 14. URL: <https://www.telegraph.co.uk/technology/2019/10/14/ai-knows-resign-do-warns-boss/> [accessed 2022-08-19]
78. Oakman J, Macdonald W, Bartram T, Keegel T, Kinsman N. Workplace risk management practices to prevent musculoskeletal and mental health disorders: what are the gaps? Safety Sci 2018 Jan;101:220-230 [FREE Full text] [doi: [10.1016/j.ssci.2017.09.004](https://doi.org/10.1016/j.ssci.2017.09.004)]
79. Maakip I, Keegel T, Oakman J. Workstyle and musculoskeletal discomfort (MSD): exploring the influence of work culture in Malaysia. J Occup Rehabil 2015 Dec;25(4):696-706 [FREE Full text] [doi: [10.1007/s10926-015-9577-2](https://doi.org/10.1007/s10926-015-9577-2)] [Medline: [25808991](https://pubmed.ncbi.nlm.nih.gov/25808991/)]
80. Li Y, Yang X, Sun P, Qi H, Lyu S. Celeb-DF: a large-scale challenging dataset for DeepFake forensics. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020 Presented at: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); Jun 13-19, 2020; Seattle, WA, USA. [doi: [10.1109/cvpr42600.2020.00327](https://doi.org/10.1109/cvpr42600.2020.00327)]
81. Cadwalladr C, Graham-Harrison E. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The guardian. 2018 Mar 17. URL: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election> [accessed 2022-08-19]
82. Isaak J, Hanna MJ. User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer 2018 Aug;51(8):56-59. [doi: [10.1109/mc.2018.3191268](https://doi.org/10.1109/mc.2018.3191268)]
83. Ahmed W. Using Twitter as a data source: an overview of current social media research tools. The London School of Economics and Political Science. 2015 Jul 10. URL: <http://blogs.lse.ac.uk/impactofsocialsciences/2015/07/10/social-media-research-tools-overview/> [accessed 2022-08-19]
84. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to twitter content. PLoS One 2015 Sep 2;10(9):e0133505 [FREE Full text] [doi: [10.1371/journal.pone.0133505](https://doi.org/10.1371/journal.pone.0133505)] [Medline: [26332588](https://pubmed.ncbi.nlm.nih.gov/26332588/)]
85. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: [10.1371/journal.pone.0198857](https://doi.org/10.1371/journal.pone.0198857)] [Medline: [29902270](https://pubmed.ncbi.nlm.nih.gov/29902270/)]
86. IPG Media Lab. 2012 May 31. URL: <https://medium.com/ipg-media-lab/twitter-use-2012-pew-research-centers-internet-american-life-project-dbbaffad4c3d> [accessed 2022-08-19]

## Abbreviations

- AI:** artificial intelligence
- GloVe:** Global Vectors for Word Representation
- HSE:** Health and Safety Executive
- LDA:** latent Dirichlet allocation
- LFS:** Labour Force Survey
- NLP:** natural language processing
- NLTK:** Natural Language Toolkit
- VADER:** Valence Aware Dictionary for Sentiment Reasoner

*Edited by A Mavragani; submitted 01.05.21; peer-reviewed by MS Kaiser, D Huang, R Gore; comments to author 29.05.21; revised version received 03.07.22; accepted 10.08.22; published 30.09.22.*

### *Please cite as:*

Hoque Tania M, Hossain MR, Jahanara N, Andreev I, Clifton DA  
Thinking Aloud or Screaming Inside: Exploratory Study of Sentiment Around Work  
JMIR Form Res 2022;6(9):e30113  
URL: <https://formative.jmir.org/2022/9/e30113>  
doi:[10.2196/30113](https://doi.org/10.2196/30113)  
PMID:[36178712](https://pubmed.ncbi.nlm.nih.gov/36178712/)

©Marzia Hoque Tania, Md Razon Hossain, Nuzhat Jahanara, Ilya Andreev, David A Clifton. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 30.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Telehealth-Supported Decision-making Psychiatric Care for Suicidal Ideation: Longitudinal Observational Study

Erin O'Callaghan<sup>1</sup>, PhD; Nicole Mahrer<sup>2</sup>, PhD; Heather G Belanger<sup>1,3</sup>, PhD; Scott Sullivan<sup>1</sup>, BA; Christine Lee<sup>1</sup>, PhD; Carina T Gupta<sup>1</sup>, BA; Mirène Winsberg<sup>1</sup>, MD

<sup>1</sup>Brightside Health Inc, Oakland, CA, United States

<sup>2</sup>Psychology Department, University of La Verne, La Verne, CA, United States

<sup>3</sup>Department of Psychiatry and Behavioral Neurosciences, University of South Florida, Tampa, FL, United States

**Corresponding Author:**

Heather G Belanger, PhD

Brightside Health Inc

2471 Peralta St, Unit A

Oakland, CA, 94607

United States

Phone: 1 813 786 7685

Email: [heather.belanger@brightside.com](mailto:heather.belanger@brightside.com)

## Abstract

**Background:** Suicide is a leading cause of death in the United States, and suicidal ideation (SI) is a significant precursor and risk factor for suicide.

**Objective:** This study aimed to examine the impact of a telepsychiatric care platform on changes in SI over time and remission, as well as to investigate the relationship between various demographic and medical factors on SI and SI remission.

**Methods:** Participants included 8581 US-based adults (8366 in the treatment group and 215 in the control group) seeking treatment for depression, anxiety, or both. The treatment group included patients who had completed at least 12 weeks of treatment and had received a prescription for at least one psychiatric medication during the study period. Providers prescribed psychiatric medications for each patient during their first session and received regular data on participants. They also received decision support at treatment onset via the digital platform, which leveraged an empirically derived proprietary precision-prescribing algorithm to give providers real-time care guidelines. Participants in the control group consisted of individuals who completed the initial enrollment data and completed surveys at baseline and 12 weeks but did not receive care.

**Results:** Greater feelings of hopelessness, anhedonia, and feeling bad about oneself were most significantly correlated ( $r=0.24-0.37$ ) with SI at baseline. Sleep issues and feeling tired or having low energy, although significant, had lower correlations with SI ( $r=0.13-0.14$ ). In terms of demographic variables, advancing age and education were associated with less SI at baseline ( $r=-0.16$ ) and 12 weeks ( $r=-0.10$ ) but less improvement over time ( $r=-0.12$  and  $-0.11$ , respectively). Although not different at baseline, the SI expression was evident in 34.4% (74/215) of the participants in the control group and 12.32% (1031/8366) of the participants in the treatment group at 12 weeks. Although the participants in the treatment group improved over time regardless of various demographic variables, participants in the control group with less education worsened over time, after controlling for age and depression severity. A model incorporating the treatment group, age, sex, and 8-item Patient Health Questionnaire scores was 77% accurate in its classification of complete remission. Those in the treatment group were 4.3 times more likely (odds ratio 4.31, 95% CI 2.88-6.44) to have complete SI remission than those in the control group. Female participants and those with advanced education beyond high school were approximately 1.4 times more likely (odds ratio 1.38, 95% CI 1.18-1.62) to remit than their counterparts.

**Conclusions:** The results highlight the efficacy of an antidepressant intervention in reducing SI, in this case administered via a telehealth platform and with decision support, as well as the importance of considering covariates, or subpopulations, when considering SI. Further research and refinement, ideally via randomized controlled trials, are needed.

(JMIR Form Res 2022;6(9):e37746) doi:[10.2196/37746](https://doi.org/10.2196/37746)



**KEYWORDS**

telemedicine; telehealth; psychiatry; mental health; suicidal ideation; depression; anxiety; suicide; depressive disorder; digital health; eHealth; psychiatric medication; demographic; psychiatric care; decision-making; decision support

## Introduction

### Background

Suicide is a leading cause of death in the United States, claiming the lives of >47,000 people in 2019 [1]. Furthermore, the prevalence of suicidal ideation (SI) is high, with 12 million adults endorsing suicidal thoughts in 2019 [1].

Amid the current global COVID-19 pandemic, concerns arose about increases in SI and suicide, with a study suggesting a particularly heightened risk at the intersection of patient vulnerability, risk, resources, and mental health status [2]. Although suicide rates remained largely unchanged or declined in the early months of the pandemic compared with the expected levels based on the prepandemic period [3], rates among adolescents [4] and young adults have increased in the aftermath [5]. A recent survey revealed significantly elevated rates of SI in those aged 18 to 24 years, minority groups, unpaid caregivers, and essential workers [6]. Overall, these trends emphasize a critical need to better understand the predictive risks of suicide and effective mediation.

Although 90% of those who commit suicide have a psychiatric diagnosis [7], predicting who will attempt suicide is difficult. SI, defined as “thinking about, considering, or planning suicide” [8], is predictive of suicide attempts and completion [9,10]. In addition, SI is a better predictor of lifetime risk for suicide than imminent risk [11]. It is estimated that among those endorsing SI, there is a 29% conditional probability of making a suicide attempt [12]. In a large retrospective study, those with nearly daily SI were 5 to 8 times more likely to attempt suicide and 3 to 11 times more likely to die by suicide within 30 days [13]. The effects of treatment with common antidepressants on SI are mixed [14-16], with the most recent review suggesting that they are associated with a higher risk of suicide [17].

Several studies have demonstrated that depression is the most common psychiatric disorder among people who die by suicide, with an estimated 50% to 75% diagnostic prevalence in suicide cases [18,19]. In addition, anxiety disorders, particularly general anxiety disorder (GAD), may be independently associated with SI and suicide attempts [20,21]. Panic disorders and attacks are associated with an increased risk of SI and suicide attempts [22]. Rates of both anxiety and depressive disorders have increased considerably in the United States in recent years, with a notable spike between April 2020 and June 2020 compared with the same period in the previous year [23], which has been suggested as a potential result of the impact of the COVID-19 pandemic on global mental health. In addition, drug and alcohol abuse has increased during the pandemic [24], both of which are associated with an increased risk for SI and suicide [25,26].

In addition to the mental health conditions associated with an increased risk of suicide, certain physical health conditions such as chronic pain and chronic medical conditions have also been shown to be associated with increased SI and suicide attempts

[27-31]. Approximately 20% of the individuals with chronic pain endorse SI [28], while 48% of the patients with fibromyalgia endorse SI [32]. Those with >1 chronic medical conditions have similarly elevated rates of SI, with 35% of those with  $\geq 2$  conditions endorsing lifetime SI [30]. After controlling for major depression and associated symptoms, as well as various demographic factors, the presence of a chronic medical condition was associated with a 1.3-times increase in the likelihood of SI [30].

Various demographic variables have been investigated as potential risk factors for SI and suicide. In general, factors such as sex (male), ethnicity (White, American Indian, or Alaska Native individuals), education level (high school or less), and economic factors (unemployment) are associated with higher rates of suicide [33-35]. Although women are more likely to have SI, men more often die by suicide [34]. Similarly, despite a low prevalence of SI in White men aged >75 years, they have one of the highest rates of fatality by suicide [11].

It has been relatively well established that suicide has a strong association with psychiatric disorders, especially major depressive disorder, and that pharmacological and nonpharmacological methods are often indicated for patients expressing SI as part of depressive symptomology. The course of treatment may commonly include prescribing antidepressants, such as selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, more modern antidepressants such as bupropion, older tricyclic antidepressants, and monoamine oxidase inhibitor antidepressants. Although antidepressants are a common treatment route, overall, there are conflicting findings regarding whether they reduce SI or suicide, or both [17,36-40]. The literature reveals a mixed and inconsistent understanding of their therapeutic effects with respect to SI and suicide. Furthermore, some studies have suggested that antidepressants may worsen suicidality in children and young adults [41,42], although this has been disputed [43,44].

### Objective

Given the limited and inconsistent understanding of the effects of psychotropic treatment on SI, this study seeks to add to the literature by investigating the impact of psychiatric care, delivered via a telehealth platform, on SI. The objective of this study was, therefore, to examine the impact of this psychiatric care platform on SI, change in SI over time, and remission, as well as to investigate the relationship between various demographic and medical factors on SI and SI remission.

## Methods

### Participants

Participant data used in this study were obtained from a national mental health telehealth company (ie, Brightside) and consisted of 8581 US-based patients receiving psychiatric care for depression or anxiety, or both between October 2018 and April

2021 (treatment, n=8366; control, n=215). Participants were eligible if they (1) completed surveys at baseline and at 12 weeks; (2) denied any history of psychosis, schizophrenia, or bipolar I disorder; and (3) denied any history of chronic liver or kidney disease. Participants in the control group met the same criteria and signed up initially for Brightside but did not receive care. Brightside uses a free self-care product that sends emails requesting the completion of survey data over a period of 14 weeks even with no sign-up. The control group therefore consisted of individuals who completed the initial enrollment

data and completed surveys at baseline and at 12 weeks. The treatment group included individuals who engaged in treatment with Brightside for at least 12 weeks. The demographic and clinical characteristics of the 2 groups are shown in Table 1. As evident in Table 1, the treatment group had significantly greater depression severity at baseline. The control group was more likely to have a high school education or less and was more likely to be unemployed. There were no significant differences on other baseline demographic or clinical characteristics, including initial SI and suicide attempts.

**Table 1.** Demographic and clinical characteristics of the sample by group.

Characteristic	Treatment (n=8366)	Control (n=215)	P value
Suicidal ideation <sup>a</sup> (baseline), mean (SD)	0.77 (0.98)	0.80 (1.04)	.63
Patient Health Questionnaire-8 score, mean (SD)	16.92 (4.38)	16.15 (5.06)	.01
Generalized Anxiety Disorder-7 score, mean (SD)	14.81 (4.52)	14.69 (4.82)	.69
Age (years), mean (SD)	32.02 (8.70)	31.97 (10.42)	.94
Sex (female), n (%)	5928 (70.86)	122 (74.39)	.34
<b>Racial minority<sup>b</sup>, n (%)</b>			
Not White	1727 (20.64)	38 (25.33)	.16
White	6639 (79.36)	112 (74.67)	.16
Black or African American	296 (3.54)	9 (6)	.16
Asian	286 (3.42)	4 (2.67)	.16
Hispanic	671 (8.02)	16 (10.67)	.16
Other	474 (5.67)	9 (6)	.16
Education (beyond high school), n (%)	5727 (68)	78 (52)	<.001
<b>Employment status, n (%)</b>			
Full-time	5738 (68.59)	135 (63.08)	.02
Part-time	975 (11.65)	19 (8.89)	.02
Unemployed by choice	808 (9.66)	28 (13.08)	.02
Unemployed	845 (10.10)	32 (14.95)	.02
<b>Chronic medical conditions, n (%)</b>			
0	6261 (75.82)	110 (73.83)	.68
1	1738 (21.05)	32 (21.48)	.68
2	235 (2.85)	6 (4.03)	.68
3	24 (0.01)	1 (0.01)	.68
Presence of chronic pain <sup>c</sup> , n (%)	851 (10.30)	28 (13.08)	.19
Panic attacks, n (%)	5712 (69.16)	147 (69.01)	.96
History of illicit drug use, n (%)	650 (7.87)	15 (7.04)	.80
History of suicide attempts, n (%)	195 (2.36)	8 (3.72)	.18

<sup>a</sup>Suicidal ideation was measured using item 9 of the Patient Health Questionnaire-9.

<sup>b</sup>Chronic pain status was missing in 1.28% (107/8366) of the patients in the treatment group.

<sup>c</sup>Participants (65/215, 30.2%) of the control group had missing data on racial minorities, gender, and education.

## Procedure

During a patient's first session, a licensed professional prescribed psychiatric medications for the patient in the treatment group. Enrolled Brightside patients completed an

initial digital intake that included clinically validated measures of depression and anxiety, as well as questions about clinical presentation, medical history, and demographics. Brightside's proprietary precision-prescribing platform analyzes these data

points using an empirically derived algorithm to provide real-time care guidelines and clinical decision support to providers via the digital platform. Brightside platform provides this decision support via a computerized symptom cluster analysis at treatment intake. On the basis of the analysis of presenting symptom clusters, as well as decision support based on the empirical literature, treatment recommendations are provided [45]. The algorithm uses a symptom cluster approach that places patients into a clinical subtype based on reported symptoms and prior medical trial results, along with comorbid medical conditions. Treatment recommendations were made according to the clinical subtype and weighted according to importance and severity and symptoms. Over the course of treatment, patients communicated with mental health providers both asynchronously via messaging and synchronously via video or telehealth sessions. The providers were psychiatrists, psychiatric nurse practitioners, or primary care physicians. These providers had an average of 12.8 contacts with patients in the first 12 weeks of treatment, which included check-in and case reviews, messaging, and video visits. Approximately half of the patients at Brightside were enrolled in a psychotherapy program. Brightside took a measurement-based approach to track long-term outcomes by prompting patients to complete periodic assessments during treatment. As such, providers were alerted in real time when patients reported improvement, failed to improve, worsened in their condition, or experienced high-risk symptoms such as SI. Most commonly prescribed regimens were monotherapy with escitalopram (1390/8366, 16.62%), bupropion (903/8366, 10.79%), sertraline (854/8366, 10.21%), or fluoxetine (502/8366, 6%), but many patients were started on some combination of different classes, with 425 different unique beginning treatment regimens and different doses. Follow-up surveys were collected at week 12 or any time before week 16.

## Measures

The Patient Health Questionnaire (PHQ-9) is a 9-item self-report measure used to assess the severity of depressive symptoms present within the prior 2 weeks, as outlined by the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* criteria. The PHQ-9 is a tool that is easy to use, but its validity remains widely confirmed [46]. Respondents rate items on a 4-point Likert scale (0-3), and total scores range from 0 to 27, with >9 indicating mild to low symptoms and >10 indicating moderate to severe symptoms [47]. The PHQ-9 shows strong validity, demonstrating 88% sensitivity and 88% specificity for major depressive disorder [47]. There is also evidence that the PHQ-9 can be used as a measure of antidepressant response [48]. In this study, patient responses to item 9 of the PHQ-9, "How often have you been bothered by thoughts that you would be better off dead or of hurting yourself in some way?" were used as a measure of SI severity [9,49]. Item 9 of the PHQ - 9 has been shown to predict suicide attempts and suicide-related deaths in an outpatient medical population [13]. For some analyses, a PHQ-8 score was calculated to avoid criterion contamination, which included the first 8 items and excluded the SI item.

The Generalized Anxiety Disorder-7 (GAD-7) scale is a 7-item self-report instrument used to assess the severity of anxiety

symptoms present within the prior 2 weeks as outlined by *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* criteria. Respondents rate items on a 4-point Likert scale (0-3), and total scores range from 1 to 21, with indications of 1 to 4 for minimal, 5 to 9 for mild, 10 to 14 for moderate, and 15 to 21 for severe symptoms [50]. The GAD-7 has shown strong reliability and validity, demonstrating 89% sensitivity and 82% specificity for GAD [51].

Basic demographic variables such as age, education, sex, and employment status were collected at baseline. In addition, individuals were asked if they had the following chronic health conditions: asthma, cancer, Crohn disease, irritable bowel syndrome, heart condition, obesity, or diabetes. A simple count variable was created with the number of chronic medical conditions endorsed. In addition, endorsement of either chronic pain or fibromyalgia was considered a variable representing chronic pain. Respondents were asked whether they used illicit substances and whether they currently experienced panic attacks. Finally, the patients were asked if they had ever attempted suicide in the past. Some participants (65/215, 30.2%) in the control group were missing the information about race, ethnicity, sex, and education level, whereas 1.28% (107/8366) of the participants in the treatment group were missing chronic pain status.

## Data Analyses

Data analyses were performed using SPSS (version 28; IBM Corp) to assemble the patient data sample, apply inclusion and exclusion criteria, and establish baseline versus follow-up survey outcomes. Brightside maintains deidentified databases for analytics that facilitate granular insights into clinical decisions, interactions, and outcomes. Assumptions for conducting regression models were assessed using visual inspection of distributions, a scatter plot of the residuals, and variance inflation factor values, as well as by examining potential multicollinearity among predictors. For the logistic regression models, the Box-Tidwell test was used to test whether the logit transform was a linear function of the predictor. In univariate general linear modeling of baseline SI severity, checking the assumptions revealed heteroscedasticity. As such, models were run using Box-Cox transformations [52]. The omnibus statistics were presented using these transformations. In all cases, the pattern and results were the same; as such, the means presented used nontransformed values for ease of interpretation.

First, we examined zero-order correlations between item 9, the SI item, on the PHQ-9 and change over time, as well as with various demographic, clinical, and medical variables. We also examined the correlations between item 9 and the other PHQ items. Pearson or point biserial correlation coefficients were calculated based on the variables included. Univariate general linear modeling was used to explore the independent effects of demographic and clinical variables on the baseline SI severity.

We then examined the relative rates of SI at baseline and 12 weeks, as well as the percentage change over time. Chi-square analyses compared relative proportions between groups. An analysis of covariance examined the treatment effect by group, on changes in SI over time, controlling for baseline age and PHQ-8 scores. Bonferroni corrections were used in all follow-up

2-tailed *t* tests. Mixed model analyses, controlling for age and PHQ-8 scores, were used to investigate the differences in SI over time by treatment group, education level, and employment status. These variables were chosen because they significantly differed between groups at baseline.

Next, using only those who endorsed SI at baseline (a score of  $\geq 1$ ), we calculated the proportion of people who *remitted* at 12 weeks (ie, they no longer endorsed any SI) and then examined potential predictors of remission. Chi-square analyses compared relative proportions among groups. Using the control group as the reference group, logistic regression with 95% CIs was used to examine bivariate relationships between SI remission and all previously described demographic, clinical, and medical variables. We then implemented a forward stepwise regression procedure to identify the independent predictors of remission during the study period, considering all the variables that were significant in the bivariate models.

### Ethics Approval

The WCG Institutional Review Board, Ethics Committee Panel 1, approved the retrospective research analysis of clinical data obtained by Brightside as part of routine clinical care (#1308524). The data were drawn from a deidentified clinical database.

## Results

### Predictors of SI Severity at Baseline

Correlations between the SI item on the PHQ-9 and all other items were examined (Table 2). Greater feelings of hopelessness, anhedonia, and feeling bad about oneself were most significantly correlated ( $r=0.24-0.37$ ) with SI at baseline. Sleep issues and feeling tired or having low energy, although significant, had

lower correlations with SI ( $r=0.13-0.14$ ). In terms of demographic variables, advancing age and education were associated with less SI at baseline ( $r=-0.16$ ) and 12 weeks ( $r=-0.10$ ), but less improvement over time ( $r=-0.12$  and  $-0.11$ , respectively). Although correlations between minority status ( $r=0.07$ ) and employment status ( $r=0.05$ ) and SI were significant, their magnitudes were small. Both the PHQ-8 and, to a lesser extent, the GAD-7 were significantly associated with SI (much more so at baseline,  $r=0.38$  and  $0.17$ , respectively) and with change over time ( $r=0.31$  and  $0.14$ , respectively). The number of chronic medical conditions was not significantly associated with SI or change over time. Endorsing either chronic pain or fibromyalgia was significantly associated with SI at baseline and week 12 but did not change over time, although the correlations were quite small ( $r=0.03-0.04$ ). Having recent panic attacks and histories of illicit drug use or suicide attempt were significantly associated with SI at both time points, as well as change over time, although only correlations with baseline SI were of any magnitude ( $r=0.08-0.10$ ).

A univariate general linear modeling was used to explore the independent effects of these demographic and clinical variables on baseline SI severity (Table 3). As GAD was highly correlated with PHQ-8 ( $r=0.35$ ;  $P<.001$ ), it was not included in the model. Similarly, having a comorbid panic disorder was correlated with sex ( $r=0.11$ ;  $P<.001$ ) and age ( $r=0.16$ ;  $P<.001$ ), so it was also excluded. The overall model was significant ( $F_{12,7733}=132.11$ ;  $P<.001$ ) and accounted for 17% of the variance in the baseline SI. As shown in Table 3, the baseline PHQ-8 score was the most prominent predictor, controlling for all other variables, and accounted for 12% of the variance in baseline SI. Other predictors, while statistically significant, did not account for much of the variance; both age and education accounted for 2% of the variance.

**Table 2.** Zero-order correlations between suicidal ideation (SI) and demographic, medical, and clinical variables (N=8581).

	SI at baseline	SI at 12 weeks	Change in SI <sup>a</sup>
<b>Demographic data</b>			
Age	-0.16 <sup>b</sup>	-0.10 <sup>b</sup>	-0.12 <sup>b</sup>
Sex (female)	-0.00	-0.03 <sup>c</sup>	0.01
Racial minority	0.07 <sup>b</sup>	-0.03 <sup>c</sup>	0.04 <sup>b</sup>
Education (beyond high school)	-0.16 <sup>b</sup>	-0.10 <sup>b</sup>	-0.11 <sup>b</sup>
Employment (less than full-time)	0.05 <sup>b</sup>	0.05 <sup>b</sup>	0.03 <sup>b</sup>
<b>PHQ<sup>d</sup> items</b>			
Anhedonia	0.24 <sup>a</sup>	0.12 <sup>a</sup>	0.18 <sup>a</sup>
Feeling down, depressed, or hopeless	0.37 <sup>a</sup>	0.15 <sup>a</sup>	0.30 <sup>a</sup>
Sleep issues	0.13 <sup>a</sup>	0.07 <sup>a</sup>	0.10 <sup>a</sup>
Tired or low energy	0.14 <sup>a</sup>	0.06 <sup>a</sup>	0.11 <sup>a</sup>
Appetite issues	0.18 <sup>a</sup>	0.09 <sup>a</sup>	0.14 <sup>a</sup>
Feeling bad about self	0.36 <sup>a</sup>	0.14 <sup>a</sup>	0.30 <sup>a</sup>
Trouble concentrating	0.18 <sup>a</sup>	0.08 <sup>a</sup>	0.15 <sup>a</sup>
Psychomotor retardation or restless	0.23 <sup>a</sup>	0.09 <sup>a</sup>	0.19 <sup>a</sup>
<b>Baseline clinical factors</b>			
PHQ-8	0.38 <sup>a</sup>	0.14 <sup>a</sup>	0.31 <sup>a</sup>
Generalized Anxiety Disorder-7	0.17 <sup>a</sup>	0.06 <sup>a</sup>	0.14 <sup>a</sup>
<b>Medical factors</b>			
Number of chronic medical conditions	0.01	0.02	-0.01
Chronic pain or fibromyalgia	0.03 <sup>b</sup>	0.04 <sup>a</sup>	0.01
Panic attacks	0.08 <sup>a</sup>	0.06 <sup>a</sup>	0.07 <sup>a</sup>
History of illicit drug use	0.08 <sup>a</sup>	0.04 <sup>a</sup>	0.05 <sup>a</sup>
History of suicide attempts	0.10 <sup>a</sup>	0.04 <sup>a</sup>	0.04 <sup>a</sup>


<sup>a</sup>Changes in SI represent changes from baseline to 12 weeks, with higher numbers representing decreased symptom severity over time on the PHQ-9 SI item.

<sup>b</sup> $P < .001$ .

<sup>c</sup> $P < .05$ .

<sup>d</sup>PHQ: Patient Health Questionnaire. PHQ-9 items are not the exact wording of the item. SI was assessed by responses to item 9 of the PHQ-9.

**Table 3.** Independent effects of demographic and clinical characteristics on baseline suicidal ideation (N=8581)<sup>a</sup>.

Factor	Sum squares	F (df)	P value	 2
Age	38.06	107.54 (1,7733)	<.001	0.02
Sex	0.21	0.60 (1,7733)	.44	0.00
Racial minority	2.44	6.90 (1,7733)	.01	0.00
Education	18.14	51.24 (1,7733)	<.001	0.02
Employment	0.56	1.51 (1,7733)	.22	0.00
Baseline PHQ-8 <sup>b</sup>	373.39	1054.98 (1,7733)	<.001	0.12
Chronic pain	0.59	1.68 (1,7733)	.20	0.00
Number of medical conditions	1.06	1.00 (1,7733)	.39	0.00
Illicit drug use	8.12	22.94 (1,7733)	<.001	0.00
Suicide attempt	5.60	15.82 (1,7733)	<.001	0.00

<sup>a</sup>The dependent measure, baseline SI score, was transformed using the Box-Cox correction.

<sup>b</sup>PHQ-8: Patient Health Questionnaire-8.

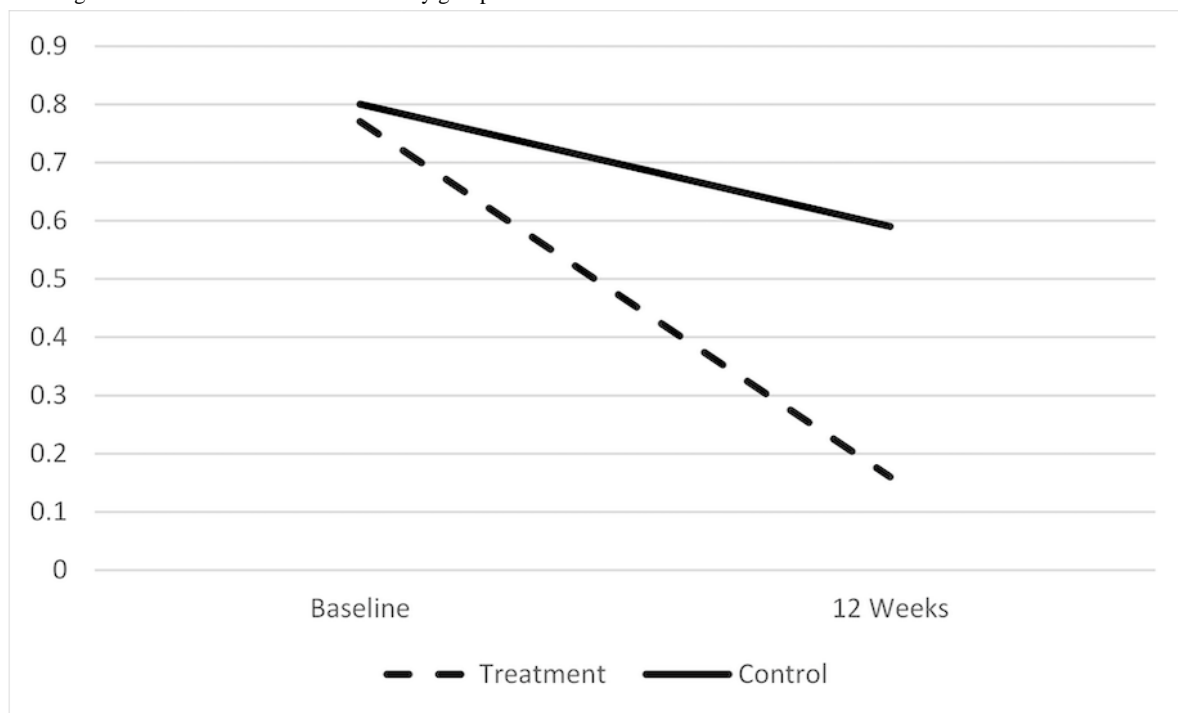
**Treatment Effects on SI Severity**

At baseline, 46.5% (100/215) of the participants in the control group and 47.12% (3942/8366) of the participants in the treatment group expressed SI ( $\chi^2_1=0.0$ ;  $P=.89$ ). At 12 weeks, 34.4% (74/215) of the participants in the control group and 12.32% (1031/8366) of the participants in the treatment group expressed SI ( $\chi^2_1=91.2$ ;  $P<.001$ ). Similarly, the percent change in SI scores was greater in the treatment group ( $F_{1,8460}=43.60$ ;  $P<.001$ ; mean 37.54%, SD 52.55%) than that in the control group (mean 13.84%, SD 66.05%); 41.73% (3491/8366) of the participants in the treatment group had lessening of SI over time, compared with 27% (58/215) of the participants in the

control group. Among the control group, 13.5% (29/215) of the participants expressed more severe SI at 12 weeks than at baseline, whereas only 2.53% (212/8366) of the participants in the treatment group did ( $\chi^2_1=106.6$ ;  $P<.001$ ).

In terms of emergence of SI in those who did not initially endorse it (n=4539), the control group had 15.6% (18/115) emergence, whereas the treatment group had only 2.98% (132/4424;  $\chi^2_1=56.2$ ;  $P<.001$ ). Analysis of covariance, controlling for age and PHQ-8 scores, demonstrated a significant effect for the treatment group ( $F_{1,7809}=145.46$ ;  $P<.001$ ) over time. **Figure 1** illustrates the change in SI over time by group, with the treatment group clearly showing significantly greater improvement (ie, reduction of SI) over time.

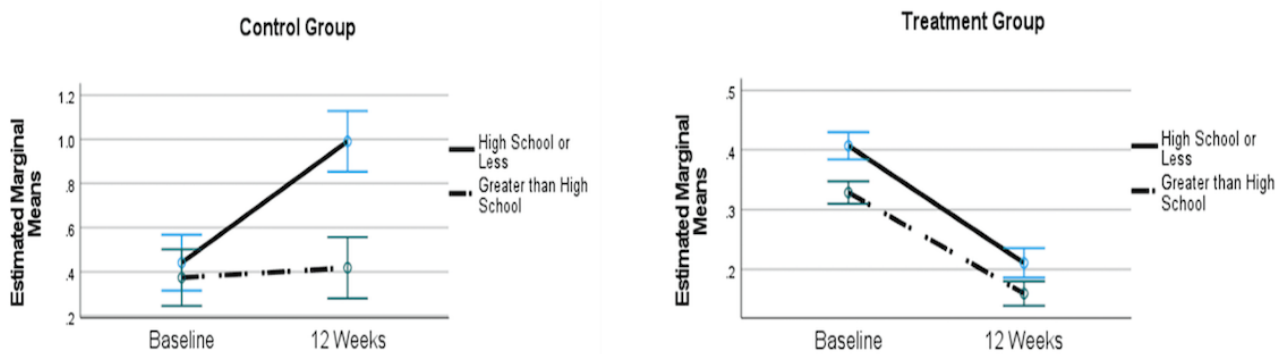
**Figure 1.** Average suicidal ideation scores over time by group.



Mixed model analyses, controlling for age and PHQ-8, investigating differences in SI over time by treatment group, education level, and employment status revealed significant effects for the treatment group ( $F_{1,7740}=46.85$ ;  $P<.001$ ), education level ( $F_{1,7740}=19.97$ ;  $P<.001$ ), and employment status ( $F_{1,7740}=6.96$ ;  $P=.01$ ). There were significant 2-way interactions, but these were further refined by a 3-way interaction among

time, group, and education level ( $F_{1,7740}=24.92$ ;  $P<.001$ ). Figure 2 shows that in the treatment group, those with both higher and lower levels of education reported less SI over time, but in the control group, education level interacted with time such that those with a high school education or less reported more SI over time (ie, got worse), whereas those with an above high school education level did not (and remained the same over time).

**Figure 2.** Interaction among group, time, and education level. Covariates appearing in the model were evaluated at age=32.03 years and PHQ-8 score=16.87. Error bars represent 95% CIs.



### Remission of SI

Among those who endorsed SI at baseline but then endorsed none at 12 weeks, 76.37% (3087/4042) of the total sample manifested this complete remission. Complete remission was observed in 77.19% (3043/3942) of the treatment group and 44% (44/100) of the control group ( $\chi^2_1=59.5$ ;  $P<.001$ ).

On the basis of bivariate relationships, the following factors were significantly associated with SI remission: being in the treatment group, older age, being female, being White (as opposed to a racial minority), obtaining education beyond high

school, and lower depression severity (as measured by PHQ-8 scores) at baseline (Table 4). Those in the treatment group were 4.3 times more likely to have SI remission than those in the control group (odds ratio [OR] 4.31, 95% CI 2.88-6.44). Women were approximately 1.4 times more likely to remit than men (OR 1.38, 95% CI 1.18-1.62). The variables that were significant in the bivariate models were entered into a binary logistic regression predicting remission using forward stepwise progression to determine the best predictors of remission, collectively. This model was 77% accurate in its classification. Table 5 shows that the treatment group, age, sex, and PHQ-8 scores were retained in the final model.

**Table 4.** Factors predicting suicidal ideation (SI) remission.

Predictor	No SI remission (n=955)	SI remission (n=3087)	Odds ratio (95% CI)
<b>Treatment, n (%)</b>			4.31 (2.88-6.44) <sup>a</sup>
Treatment group (n=3942)	899 (22.81)	3043 (77.19)	
Control group (Ref <sup>b</sup> ) (n=100)	56 (56)	44 (44)	
<b>Demographics</b>			
Age (years), mean (SD)	29.43 (8.20)	30.81 (8.28)	1.02 (1.01-1.03) <sup>a</sup>
<b>Sex, n (%)</b>			1.38 (1.18-1.62) <sup>a</sup>
Female (n=2875)	626 (21.77)	2249 (78.23)	
Male (n=1144)	318 (27.79)	826 (72.2)	
<b>Minority status, n (%)</b>			0.79 (0.67-0.94) <sup>a</sup>
Minority (n=934)	250 (26.77)	684 (73.23)	
White (Ref) (n=3079)	690 (22.41)	2389 (77.59)	
<b>Education, n (%)</b>			1.37 (1.18-1.59) <sup>a</sup>
Beyond high school (n=2466)	523 (21.21)	1943 (78.79)	
High school or less (Ref) (n=1547)	417 (26.96)	1130 (73.04)	
<b>Employment, n (%)</b>			0.95 (0.76-1.19)
Unemployed (n=466)	114 (24.46)	352 (75.54)	
Employed (Ref) (n=3575)	840 (23.5)	2735 (76.5)	
<b>Clinical, mean (SD)</b>			
PHQ-8 <sup>c</sup>	21.01 (3.98)	19.81 (4.10)	0.94 (0.92-0.96) <sup>a</sup>
GAD-7 <sup>d</sup>	15.57 (4.42)	15.28 (4.46)	0.99 (0.97-1.00)
<b>Medical condition, n (%)</b>			
<b>Chronic pain or fibromyalgia</b>			0.81 (0.65-1.02)
Chronic pain (n=456)	123 (26.97)	333 (73.03)	
No chronic pain (Ref) (n=3537)	818 (23.13)	2719 (76.87)	
<b>Number of chronic medical issues</b>			0.91 (0.79-1.05)
0 (Ref) (n=2993)	695 (23.22)	2298 (76.78)	
1 (n=853)	193 (22.63)	660 (77.37)	
2 (n=106)	34 (32.08)	72 (67.92)	
3 (n=12)	5 (41.67)	7 (58.33)	
<b>Panic attacks</b>			0.79 (0.67-0.94)
Yes (n=2939)	724 (24.63)	2215 (75.37)	
No (Ref) (n=1053)	217 (20.61)	836 (79.39)	
<b>Illicit drug use</b>			0.86 (0.68-1.08)
Yes (n=405)	106 (26.17)	299 (73.83)	
No (Ref) (n=3587)	835 (23.28)	2752 (76.72)	
<b>Suicide attempts</b>			0.85 (0.57-1.27)
Yes (n=128)	34 (26.56)	94 (73.44)	
No (Ref) (n=3867)	908 (23.48)	2959 (76.52)	

<sup>a</sup>Values indicate that the predictor significantly predicts SI remission at the 95% CI.

<sup>b</sup>Ref. represents the reference group.

<sup>c</sup>PHQ-8: Patient Health Questionnaire-8.



<sup>d</sup>GAD-7: Generalized Anxiety Disorder-7.

**Table 5.** Multivariate model predicting suicidal ideation remission.

Factor	Odds ratio <sup>a</sup> (95% CI)
Treatment group	5.10 (3.03-8.43)
Age	1.02 (1.01-1.03)
Sex	1.49 (1.26-1.77)
PHQ-8 <sup>b</sup>	0.94 (0.92-0.96)

<sup>a</sup>All odds ratios are  $P < .001$ ; Nagelkerke  $R^2 = 0.04$ .

<sup>b</sup>PHQ-8: Patient Health Questionnaire-8.

## Discussion

The objective of this study was to examine the impact of psychiatric care on SI, change in SI over time, and remission, as well as to investigate the relationship between various demographic and medical factors on SI and SI remission.

### SI Severity

Greater feelings of hopelessness, anhedonia, and feeling bad about oneself were most significantly correlated with SI at baseline. Sleep issues and feeling tired or having low energy, although significant, had lower correlations with SI. These patterns of associations between the SI item and other items of the PHQ-9 mirror those found in other studies with primary care patients who had depression or chronic pain and had completed the PHQ-9 [53], with the exception of anhedonia, which tended to have lower associations with SI in primary care patients [53]. Hopelessness has consistently been found to be a predictor of SI [54,55], although associations with actual suicide or attempts are mixed [56-58].

Associations between greater SI severity and younger age are consistent with national survey data finding that younger adults more frequently endorse SI than older adults [59]. As might be expected, educational levels beyond high school were associated with lower SI. The findings that advanced education and age were associated with less positive change in SI are difficult to reconcile though this may be because of range restriction (ie, there is less room to “improve”). In the Sequenced Treatment Alternatives to Relieve Depression (STAR\*D) data, greater educational level and older age predicted improvement or lowering of SI [16].

Not surprisingly, both the PHQ-8 and, to a lesser extent, the GAD-7 were significantly associated with SI (much more so at baseline) and with change over time. The baseline PHQ-8 score was the most prominent independent predictor of SI, accounting for 12% of the variance. This reflects a consistent finding in the literature that depression severity is highly associated with SI [60] and with suicide-related outcomes [61] and illustrates the importance of antidepressant treatment.

The number of chronic medical conditions was not significantly associated with SI or change over time. Endorsing either chronic pain or fibromyalgia was significantly associated with SI at baseline and week 12 but did not change over time, although the overall correlations were quite small. This is counter to prior

research [62,63] and may reflect the limited range (the maximum number of endorsed medical conditions was 3) or failure to consider the severity or burden of the medical conditions. Having recent panic attacks and histories of illicit drug use or suicide attempt were significantly associated with SI at both time points, as well as change over time, although only correlations with baseline SI were of any magnitude ( $r = 0.08-0.10$ ). Again, these findings are in line with prior research [22,64-66], although it is important to note that approximately 11% of those who attempt suicide deny ever having experienced any SI [66]. Therefore, SI is insufficient to explain all suicide attempts and completions.

### Treatment Effects

Although the groups were similar at baseline in terms of the presence of SI (47%) at baseline, after 12 weeks of treatment, only 12.32% (1031/8366) of the participants in the treatment group expressed any SI, compared with 34% (74/215) of the participants in the control group. These numbers are similar to a much smaller study investigating psychotherapy for depression [67], although that study's participants had fewer people (30%) with SI at baseline than this study and less severe depression. In this study, 41.73% (3491/8366) of the participants in the treatment group showed improvement in SI severity over time, compared with 27% (58/215) of the participants in the control group. The baseline rate of SI and the percentage of people who were improved at the last visit were very similar to those in the STAR\*D trial [68], which was conducted in person, as compared with this study conducted virtually. In terms of the emergence of new SI in those not initially endorsing it, 15.6% (18/115) of the participants in the control group showed emergent SI, compared with only 2.98% (132/4424) of the participants in the treatment group. These proportions are similar to the 1.3% emergence seen in the STAR\*D trial at 12 weeks [67,68], although the STAR\*D trial was treated in a more homogenous fashion in person. Further research is needed to replicate these promising treatment effects with virtual treatment and clinical decision support tools for more tailored precision prescribing.

Although the treatment group improved over time regardless of various demographic variables, in the control group, after controlling for age and depression severity, those with less education worsened over time. Greater education levels are protective against many adverse outcomes, including SI [69-71]. Telehealth interventions, such as the Brightside platform used

with the treatment group in this study, may be indicated for those at risk because of lower education.

### Remission

Those in the treatment group were 4.3 times more likely to remit than those in the control group (OR 4.31, 95% CI 2.88-6.44). Zisook et al [40] found similar rates of remission in those receiving 12 weeks of escitalopram plus placebo, bupropion sustained release plus escitalopram or venlafaxine extended release plus mirtazapine. Collectively, these findings suggest strong evidence that antidepressant medications have a positive impact on SI. Whether selective serotonin reuptake inhibitors and other new-generation antidepressants alter the risk of suicide in adults is certainly controversial [38,39,72,73], with the most recent review suggesting that they are associated with a higher risk of suicide [17]. These authors contend that publication bias and conflicts of interest likely contribute to systematic underestimation of risk. Although this study involved SI and did not evaluate suicide attempts or completion to the extent that SI is predictive of suicide attempts and actual suicide [9,10], these data are more optimistic.

Treatment was clearly the biggest predictor of SI remission. Other factors found to be significantly associated with SI remission were older age, being female, being White, obtaining education beyond high school, and having lower depression severity at baseline. Women and those with advanced education beyond high school were about 1.4 times (OR 1.38, 95% CI 1.18-1.62) more likely to remit than men and those without advanced education (OR 1.37, 95% CI 1.18-1.59). In the STAR\*D trial, remission of depressive symptoms was more likely in those who were White, female, employed, or had higher levels of education or income [74]. Although not specifically focused on SI, these factors are similar to those found in this study.

### Limitations

The primary limitation of this study is that there was no random assignment to treatment, and alternative explanations for any observed treatment effects are possible. Although comparison to a control group that completed assessments on the same schedule as the treatment group, and which was largely equivalent to the treatment group at baseline, reduced the likelihood that any effects were because of engagement on line or other Hawthorne effects, the control group did not engage with providers. As the participants were not randomized, there is potential for confounding. For example, participants in the control group were more likely to be unemployed and have no

education beyond high school. As it is unclear why the control group participants did not pursue treatment, one possibility is that they had less scheduling flexibility or less ability to take time off to attend treatment.

Another limitation is the inability to directly compare different medications, as this was a clinical sample with >400 different combinations of medications. Although this prevents generalizability to specific medication groups, it does speak to the ability of antidepressant treatment, as rendered in this novel virtual treatment regimen, to positively affect SI. An additional limitation, however, is that patients may have had other treatments such as psychotherapy that were not assessed. In addition, a specific measure of SI, such as the Beck Scale for Suicide Ideation or the Scale for Suicide Ideation [75], would have improved the study rather than a single item from a global measure of depression. Further research will be required to study the impact of this prescription model on suicide attempt.

### Treatment Implications and Conclusions

Depression severity is the primary driver of SI, relative to demographic and other clinical or medical factors. Certainly, many clinicians may be reluctant to prescribe antidepressants in those with SI because of the perceived risk of working with patients who are suicidal. These results address, at least to some degree, these concerns. The results of this study, as well as those of others, are consistent with the efficacy of psychiatric care administered via a telehealth platform with decision support. In antidepressant trials, depression severity mediates the effect of antidepressant medication on suicide risk [76,77], so treating the depressive symptoms is paramount. In this study and others, antidepressant medication positively affected SI severity in the vast majority of people. Finally, additional efforts should be made to treat those with lower levels of education, as they are more likely to have increasing SI over time without such treatment. These findings highlight the importance of considering covariates, or subpopulations, when considering SI [78].

Finally, these results align with a growing body of literature demonstrating the effectiveness of using a telehealth platform for providing mental health services [79-82]. Clinical decision support tools have demonstrated efficacy in depression outcomes in primary care and general practices [83-85]. The strengths of this study include its size, use of a control group, and the novel use of a telehealth platform. Further research and refinement, ideally via randomized controlled trials, are needed.

### Conflicts of Interest

EOC, HGB, SS, and MW hold stocks in Brightside Health Inc.

### References

1. Suicide statistics New York, NY2021. American Foundation for Suicide Prevention. URL: <https://afsp.org/suicide-statistics/> [accessed 2022-02-02]
2. Fitzpatrick KM, Harris C, Drawve G. How bad is it? Suicidality in the middle of the COVID-19 pandemic. *Suicide Life Threat Behav* 2020 Dec;50(6):1241-1249 [FREE Full text] [doi: [10.1111/sltb.12655](https://doi.org/10.1111/sltb.12655)] [Medline: [32589799](https://pubmed.ncbi.nlm.nih.gov/32589799/)]

3. Pirkis J, John A, Shin S, DelPozo-Banos M, Arya V, Analuisa-Aguilar P, et al. Suicide trends in the early months of the COVID-19 pandemic: an interrupted time-series analysis of preliminary data from 21 countries. *Lancet Psychiatry* 2021 Jul;8(7):579-588 [FREE Full text] [doi: [10.1016/S2215-0366\(21\)00091-2](https://doi.org/10.1016/S2215-0366(21)00091-2)] [Medline: [33862016](https://pubmed.ncbi.nlm.nih.gov/33862016/)]
4. Charpignon ML, Ontiveros J, Sundaresan S, Puri A, Chandra J, Mandl KD, et al. Evaluation of suicides among US adolescents during the COVID-19 pandemic. *JAMA Pediatr* 2022 Jul 01;176(7):724-726. [doi: [10.1001/jamapediatrics.2022.0515](https://doi.org/10.1001/jamapediatrics.2022.0515)] [Medline: [35467724](https://pubmed.ncbi.nlm.nih.gov/35467724/)]
5. Clapperton A, Spittal MJ, Dwyer J, Garrett A, Kölves K, Leske S, et al. Patterns of suicide in the context of COVID-19: evidence from three Australian states. *Front Psychiatry* 2021 Nov 30;12:797601 [FREE Full text] [doi: [10.3389/fpsy.2021.797601](https://doi.org/10.3389/fpsy.2021.797601)] [Medline: [34916981](https://pubmed.ncbi.nlm.nih.gov/34916981/)]
6. Czeisler M, 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](https://doi.org/10.15585/mmwr.mm6932a1)] [Medline: [32790653](https://pubmed.ncbi.nlm.nih.gov/32790653/)]
7. Beautrais AL, Joyce PR, Mulder RT, Fergusson DM, Deavoll BJ, Nightingale SK. Prevalence and comorbidity of mental disorders in persons making serious suicide attempts: a case-control study. *Am J Psychiatry* 1996 Aug;153(8):1009-1014. [doi: [10.1176/ajp.153.8.1009](https://doi.org/10.1176/ajp.153.8.1009)] [Medline: [8678168](https://pubmed.ncbi.nlm.nih.gov/8678168/)]
8. Suicide. National Institute of Mental Health. Bethesda, MD, USA: National Institute of Mental Health; 2022. URL: <https://www.nimh.nih.gov/health/statistics/suicide> [accessed 2022-02-02]
9. Simon GE, Rutter CM, Peterson D, Oliver M, Whiteside U, Operskalski B, et al. Does response on the PHQ-9 Depression Questionnaire predict subsequent suicide attempt or suicide death? *Psychiatr Serv* 2013 Dec 01;64(12):1195-1202 [FREE Full text] [doi: [10.1176/appi.ps.201200587](https://doi.org/10.1176/appi.ps.201200587)] [Medline: [24036589](https://pubmed.ncbi.nlm.nih.gov/24036589/)]
10. Simon GE, Yarborough BJ, Rossom RC, Lawrence JM, Lynch FL, Waitzfelder BE, et al. Self-reported suicidal ideation as a predictor of suicidal behavior among outpatients with diagnoses of psychotic disorders. *Psychiatr Serv* 2019 Mar 01;70(3):176-183 [FREE Full text] [doi: [10.1176/appi.ps.201800381](https://doi.org/10.1176/appi.ps.201800381)] [Medline: [30526341](https://pubmed.ncbi.nlm.nih.gov/30526341/)]
11. Harmer B, Lee S, Duong TH, Saadabadi A. Suicidal Ideation. Treasure Island, FL, USA: StatPearls Publishing; May 18, 2022.
12. Nock MK, Stein MB, Heeringa SG, Ursano RJ, Colpe LJ, Fullerton CS, Army STARRS Collaborators. Prevalence and correlates of suicidal behavior among soldiers: results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry* 2014 May;71(5):514-522 [FREE Full text] [doi: [10.1001/jamapsychiatry.2014.30](https://doi.org/10.1001/jamapsychiatry.2014.30)] [Medline: [24590178](https://pubmed.ncbi.nlm.nih.gov/24590178/)]
13. Rossom RC, Coleman KJ, Ahmedani BK, Beck A, Johnson E, Oliver M, et al. Suicidal ideation reported on the PHQ9 and risk of suicidal behavior across age groups. *J Affect Disord* 2017 Jun;215:77-84 [FREE Full text] [doi: [10.1016/j.jad.2017.03.037](https://doi.org/10.1016/j.jad.2017.03.037)] [Medline: [28319695](https://pubmed.ncbi.nlm.nih.gov/28319695/)]
14. Stone M, Laughren T, Jones ML, Levenson M, Holland PC, Hughes A, et al. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. *BMJ* 2009 Aug 11;339:b2880 [FREE Full text] [doi: [10.1136/bmj.b2880](https://doi.org/10.1136/bmj.b2880)] [Medline: [19671933](https://pubmed.ncbi.nlm.nih.gov/19671933/)]
15. Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E, Sarchiapone M, et al. Suicide prevention strategies revisited: 10-year systematic review. *Lancet Psychiatry* 2016 Jul;3(7):646-659. [doi: [10.1016/S2215-0366\(16\)30030-X](https://doi.org/10.1016/S2215-0366(16)30030-X)] [Medline: [27289303](https://pubmed.ncbi.nlm.nih.gov/27289303/)]
16. Weissman CR, Hadas I, Yu D, Jones B, Kong D, Mulsant BH, et al. Predictors of change in suicidal ideation across treatment phases of major depressive disorder: analysis of the STAR\*D data. *Neuropsychopharmacology* 2021 Jun;46(7):1293-1299 [FREE Full text] [doi: [10.1038/s41386-020-00953-9](https://doi.org/10.1038/s41386-020-00953-9)] [Medline: [33479508](https://pubmed.ncbi.nlm.nih.gov/33479508/)]
17. Hengartner MP, Amendola S, Kaminski JA, Kindler S, Bschor T, Plöderl M. Suicide risk with selective serotonin reuptake inhibitors and other new-generation antidepressants in adults: a systematic review and meta-analysis of observational studies. *J Epidemiol Community Health* 2021 Mar 08;75(6):523-530. [doi: [10.1136/jech-2020-214611](https://doi.org/10.1136/jech-2020-214611)] [Medline: [33685964](https://pubmed.ncbi.nlm.nih.gov/33685964/)]
18. Conwell Y, Duberstein PR, Cox C, Herrmann JH, Forbes NT, Caine ED. Relationships of age and axis I diagnoses in victims of completed suicide: a psychological autopsy study. *Am J Psychiatry* 1996 Aug;153(8):1001-1008. [doi: [10.1176/ajp.153.8.1001](https://doi.org/10.1176/ajp.153.8.1001)] [Medline: [8678167](https://pubmed.ncbi.nlm.nih.gov/8678167/)]
19. Rich CL, Young D, Fowler RC. San Diego suicide study. I. Young vs old subjects. *Arch Gen Psychiatry* 1986 Jun;43(6):577-582. [doi: [10.1001/archpsyc.1986.01800060071009](https://doi.org/10.1001/archpsyc.1986.01800060071009)] [Medline: [3707290](https://pubmed.ncbi.nlm.nih.gov/3707290/)]
20. Bentley KH, Franklin JC, Ribeiro JD, Kleiman EM, Fox KR, Nock MK. Anxiety and its disorders as risk factors for suicidal thoughts and behaviors: a meta-analytic review. *Clin Psychol Rev* 2016 Feb;43:30-46 [FREE Full text] [doi: [10.1016/j.cpr.2015.11.008](https://doi.org/10.1016/j.cpr.2015.11.008)] [Medline: [26688478](https://pubmed.ncbi.nlm.nih.gov/26688478/)]
21. Thibodeau MA, Welch PG, Sareen J, Asmundson GJ. Anxiety disorders are independently associated with suicide ideation and attempts: propensity score matching in two epidemiological samples. *Depress Anxiety* 2013 Oct;30(10):947-954. [doi: [10.1002/da.22203](https://doi.org/10.1002/da.22203)] [Medline: [24108489](https://pubmed.ncbi.nlm.nih.gov/24108489/)]
22. Weissman MM, Klerman GL, Markowitz JS, Ouellette R. Suicidal ideation and suicide attempts in panic disorder and attacks. *N Engl J Med* 1989 Nov 02;321(18):1209-1214. [doi: [10.1056/NEJM198911023211801](https://doi.org/10.1056/NEJM198911023211801)] [Medline: [2797086](https://pubmed.ncbi.nlm.nih.gov/2797086/)]

23. Indicators of Anxiety or Depression Based on Reported Frequency of Symptoms During Last 7 Days. Centers for Disease Control and Prevention. Atlanta, GA, USA: US Department of Health and Human Services; 2020. URL: <https://data.cdc.gov/NCHS/Indicators-of-Anxiety-or-Depression-Based-on-Report/8pt5-q6wp> [accessed 2022-02-02]
24. Roberts A, Rogers J, Mason R, Siriwardena AN, Hogue T, Whitley GA, et al. Alcohol and other substance use during the COVID-19 pandemic: a systematic review. *Drug Alcohol Depend* 2021 Dec 01;229(Pt A):109150 [FREE Full text] [doi: [10.1016/j.drugalcdep.2021.109150](https://doi.org/10.1016/j.drugalcdep.2021.109150)] [Medline: [34749198](https://pubmed.ncbi.nlm.nih.gov/34749198/)]
25. Bragazzi NL, Beamish D, Kong JD, Wu J. Illicit drug use in Canada and implications for suicidal behaviors, and household food insecurity: findings from a large, nationally representative survey. *Int J Environ Res Public Health* 2021 Jun 14;18(12):6425 [FREE Full text] [doi: [10.3390/ijerph18126425](https://doi.org/10.3390/ijerph18126425)] [Medline: [34198517](https://pubmed.ncbi.nlm.nih.gov/34198517/)]
26. Carmel A, Ries R, West II, Bumgardner K, Roy-Byrne P. Suicide risk and associated demographic and clinical correlates among primary care patients with recent drug use. *Am J Drug Alcohol Abuse* 2016 May;42(3):351-357 [FREE Full text] [doi: [10.3109/00952990.2015.1133634](https://doi.org/10.3109/00952990.2015.1133634)] [Medline: [26910262](https://pubmed.ncbi.nlm.nih.gov/26910262/)]
27. Fishbain DA, Lewis JE, Gao J. The pain suicidality association: a narrative review. *Pain Med* 2014 Nov;15(11):1835-1849. [doi: [10.1111/pme.12463](https://doi.org/10.1111/pme.12463)] [Medline: [24995953](https://pubmed.ncbi.nlm.nih.gov/24995953/)]
28. Tang NK, Crane C. Suicidality in chronic pain: a review of the prevalence, risk factors and psychological links. *Psychol Med* 2006 May;36(5):575-586. [doi: [10.1017/S0033291705006859](https://doi.org/10.1017/S0033291705006859)] [Medline: [16420727](https://pubmed.ncbi.nlm.nih.gov/16420727/)]
29. Chiu LW, Lin KY, Chang CY, Lin WC, Wan L. Increased suicide risk among patients oophorectomized following benign conditions and its association with comorbidities. *J Psychosom Obstet Gynaecol* 2020 Jun;41(2):137-143. [doi: [10.1080/0167482X.2019.1658735](https://doi.org/10.1080/0167482X.2019.1658735)] [Medline: [31462149](https://pubmed.ncbi.nlm.nih.gov/31462149/)]
30. Druss B, Pincus H. Suicidal ideation and suicide attempts in general medical illnesses. *Arch Intern Med* 2000 May 22;160(10):1522-1526. [doi: [10.1001/archinte.160.10.1522](https://doi.org/10.1001/archinte.160.10.1522)] [Medline: [10826468](https://pubmed.ncbi.nlm.nih.gov/10826468/)]
31. Alias A, Bertrand L, Bisson-Gervais V, Henry M. Suicide in obstructive lung, cardiovascular and oncological disease. *Prev Med* 2021 Nov;152(Pt 1):106543. [doi: [10.1016/j.ypmed.2021.106543](https://doi.org/10.1016/j.ypmed.2021.106543)] [Medline: [34538370](https://pubmed.ncbi.nlm.nih.gov/34538370/)]
32. Calandre EP, Navajas-Rojas MA, Ballesteros J, Garcia-Carrillo J, Garcia-Leiva JM, Rico-Villademoros F. Suicidal ideation in patients with fibromyalgia: a cross-sectional study. *Pain Pract* 2015 Feb;15(2):168-174. [doi: [10.1111/papr.12164](https://doi.org/10.1111/papr.12164)] [Medline: [24433278](https://pubmed.ncbi.nlm.nih.gov/24433278/)]
33. Stack S. Suicide: a 15-year review of the sociological literature. Part I: cultural and economic factors. *Suicide Life Threat Behav* 2000;30(2):145-162. [Medline: [10888055](https://pubmed.ncbi.nlm.nih.gov/10888055/)]
34. Suicide Prevention. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/suicide/index.html> [accessed 2022-02-02]
35. Phillips JA, Hempstead K. Differences in U.S. suicide rates by educational attainment, 2000-2014. *Am J Prev Med* 2017 Oct;53(4):e123-e130. [doi: [10.1016/j.amepre.2017.04.010](https://doi.org/10.1016/j.amepre.2017.04.010)] [Medline: [28756896](https://pubmed.ncbi.nlm.nih.gov/28756896/)]
36. Khan A, Khan S, Kolts R, Brown WA. Suicide rates in clinical trials of SSRIs, other antidepressants, and placebo: analysis of FDA reports. *Am J Psychiatry* 2003 Apr;160(4):790-792. [doi: [10.1176/appi.ajp.160.4.790](https://doi.org/10.1176/appi.ajp.160.4.790)] [Medline: [12668373](https://pubmed.ncbi.nlm.nih.gov/12668373/)]
37. Tondo L, Baldessarini RJ. Suicidal behavior in mood disorders: response to pharmacological treatment. *Curr Psychiatry Rep* 2016 Sep;18(9):88. [doi: [10.1007/s11920-016-0715-0](https://doi.org/10.1007/s11920-016-0715-0)] [Medline: [27542851](https://pubmed.ncbi.nlm.nih.gov/27542851/)]
38. Reeves RR, Ladner ME. Antidepressant-induced suicidality: an update. *CNS Neurosci Ther* 2010 Aug;16(4):227-234 [FREE Full text] [doi: [10.1111/j.1755-5949.2010.00160.x](https://doi.org/10.1111/j.1755-5949.2010.00160.x)] [Medline: [20553304](https://pubmed.ncbi.nlm.nih.gov/20553304/)]
39. Kaminski JA, Bschor T. Antidepressants and suicidality: a re-analysis of the re-analysis. *J Affect Disord* 2020 Apr 01;266:95-99. [doi: [10.1016/j.jad.2020.01.107](https://doi.org/10.1016/j.jad.2020.01.107)] [Medline: [32056952](https://pubmed.ncbi.nlm.nih.gov/32056952/)]
40. Zisook S, Lesser IM, Lebowitz B, Rush AJ, Kallenberg G, Wisniewski SR, et al. Effect of antidepressant medication treatment on suicidal ideation and behavior in a randomized trial: an exploratory report from the Combining Medications to Enhance Depression Outcomes Study. *J Clin Psychiatry* 2011 Oct;72(10):1322-1332. [doi: [10.4088/JCP.10m06724](https://doi.org/10.4088/JCP.10m06724)] [Medline: [22075098](https://pubmed.ncbi.nlm.nih.gov/22075098/)]
41. Barbui C, Esposito E, Cipriani A. Selective serotonin reuptake inhibitors and risk of suicide: a systematic review of observational studies. *CMAJ* 2009 Feb 03;180(3):291-297 [FREE Full text] [doi: [10.1503/cmaj.081514](https://doi.org/10.1503/cmaj.081514)] [Medline: [19188627](https://pubmed.ncbi.nlm.nih.gov/19188627/)]
42. Hammad TA, Laughren TP, Racoosin JA. Suicide rates in short-term randomized controlled trials of newer antidepressants. *J Clin Psychopharmacol* 2006 Apr;26(2):203-207. [doi: [10.1097/01.jcp.0000203198.11453.95](https://doi.org/10.1097/01.jcp.0000203198.11453.95)] [Medline: [16633153](https://pubmed.ncbi.nlm.nih.gov/16633153/)]
43. Kaizar EE, Greenhouse JB, Seltman H, Kelleher K. Do antidepressants cause suicidality in children? A Bayesian meta-analysis. *Clin Trials* 2006;3(2):73-98. [doi: [10.1191/1740774506cn139oa](https://doi.org/10.1191/1740774506cn139oa)] [Medline: [16773951](https://pubmed.ncbi.nlm.nih.gov/16773951/)]
44. Gibbons RD, Hur K, Bhaumik DK, Mann JJ. The relationship between antidepressant medication use and rate of suicide. *Arch Gen Psychiatry* 2005 Feb;62(2):165-172. [doi: [10.1001/archpsyc.62.2.165](https://doi.org/10.1001/archpsyc.62.2.165)] [Medline: [15699293](https://pubmed.ncbi.nlm.nih.gov/15699293/)]
45. O'Callaghan E, Sullivan S, Gupta C, Belanger HG, Winsberg M. Feasibility and acceptability of a novel telepsychiatry-delivered precision prescribing intervention for anxiety and depression. *BMC Psychiatry* 2022 Jul 19;22(1):483 [FREE Full text] [doi: [10.1186/s12888-022-04113-9](https://doi.org/10.1186/s12888-022-04113-9)] [Medline: [35854281](https://pubmed.ncbi.nlm.nih.gov/35854281/)]
46. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. *J Affect Disord* 2021 Jan 15;279:473-483. [doi: [10.1016/j.jad.2020.09.131](https://doi.org/10.1016/j.jad.2020.09.131)] [Medline: [33126078](https://pubmed.ncbi.nlm.nih.gov/33126078/)]

47. 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](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
48. Löwe B, Schenkel I, Carney-Doebbeling C, Göbel C. Responsiveness of the PHQ-9 to psychopharmacological depression treatment. *Psychosomatics* 2006;47(1):62-67. [doi: [10.1176/appi.psy.47.1.62](https://doi.org/10.1176/appi.psy.47.1.62)] [Medline: [16384809](https://pubmed.ncbi.nlm.nih.gov/16384809/)]
49. Yarborough BJ, Stumbo SP, Ahmedani B, Rossom R, Coleman K, Boggs JM, et al. Suicide behavior following PHQ-9 screening among individuals with substance use disorders. *J Addict Med* 2021;15(1):55-60. [doi: [10.1097/ADM.0000000000000696](https://doi.org/10.1097/ADM.0000000000000696)] [Medline: [32657957](https://pubmed.ncbi.nlm.nih.gov/32657957/)]
50. 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](https://doi.org/10.1016/j.genhosppsych.2015.11.005)] [Medline: [26719105](https://pubmed.ncbi.nlm.nih.gov/26719105/)]
51. 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](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]
52. Osborne J. Improving your data transformations: applying the Box-Cox transformation. *Pract Assess Res Eval* 2010;15(12):1-9 [FREE Full text] [doi: [10.7275/qbpc-gk17](https://doi.org/10.7275/qbpc-gk17)]
53. Kroenke K, Wu J, Yu Z, Bair MJ, Kean J, Stump T, et al. Patient health questionnaire anxiety and depression scale: initial validation in three clinical trials. *Psychosom Med* 2016;78(6):716-727 [FREE Full text] [doi: [10.1097/PSY.0000000000000322](https://doi.org/10.1097/PSY.0000000000000322)] [Medline: [27187854](https://pubmed.ncbi.nlm.nih.gov/27187854/)]
54. Brown GK, Beck AT, Steer RA, Grisham JR. Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. *J Consult Clin Psychol* 2000 Jun;68(3):371-377. [Medline: [10883553](https://pubmed.ncbi.nlm.nih.gov/10883553/)]
55. Isometsä E. Suicidal behaviour in mood disorders--who, when, and why? *Can J Psychiatry* 2014 Mar;59(3):120-130 [FREE Full text] [doi: [10.1177/070674371405900303](https://doi.org/10.1177/070674371405900303)] [Medline: [24881160](https://pubmed.ncbi.nlm.nih.gov/24881160/)]
56. Oquendo MA, Currier D, Mann JJ. Prospective studies of suicidal behavior in major depressive and bipolar disorders: what is the evidence for predictive risk factors? *Acta Psychiatr Scand* 2006 Sep;114(3):151-158. [doi: [10.1111/j.1600-0447.2006.00829.x](https://doi.org/10.1111/j.1600-0447.2006.00829.x)] [Medline: [16889585](https://pubmed.ncbi.nlm.nih.gov/16889585/)]
57. Samuelsson M, Jokinen J, Nordström AL, Nordström P. CSF 5-HIAA, suicide intent and hopelessness in the prediction of early suicide in male high-risk suicide attempters. *Acta Psychiatr Scand* 2006 Jan;113(1):44-47. [doi: [10.1111/j.1600-0447.2005.00639.x](https://doi.org/10.1111/j.1600-0447.2005.00639.x)] [Medline: [16390368](https://pubmed.ncbi.nlm.nih.gov/16390368/)]
58. Young MA, Fogg LF, Scheffner WA, Fawcett JA. Interactions of risk factors in predicting suicide. *Am J Psychiatry* 1994 Mar;151(3):434-435. [doi: [10.1176/ajp.151.3.434](https://doi.org/10.1176/ajp.151.3.434)] [Medline: [8109656](https://pubmed.ncbi.nlm.nih.gov/8109656/)]
59. Behavioral Health Barometer: United States, Volume 5: Indicators as measured through the 2017 National Survey on Drug Use and Health and the National Survey of Substance Abuse Treatment Services. Rockville, MD, USA: Substance Abuse and Mental Health Services Administration (US); 2019.
60. Macalli M, Navarro M, Orri M, Tournier M, Thiébaud R, Côté SM, et al. A machine learning approach for predicting suicidal thoughts and behaviours among college students. *Sci Rep* 2021 Jun 15;11(1):11363 [FREE Full text] [doi: [10.1038/s41598-021-90728-z](https://doi.org/10.1038/s41598-021-90728-z)] [Medline: [34131161](https://pubmed.ncbi.nlm.nih.gov/34131161/)]
61. Fiedorowicz JG, Persons JE, Assari S, Ostacher MJ, Goes FS, Nurnberger JI, et al. Moderators of the association between depressive, manic, and mixed mood symptoms and suicidal ideation and behavior: an analysis of the National Network of Depression Centers Mood Outcomes Program. *J Affect Disord* 2021 Feb 15;281:623-630 [FREE Full text] [doi: [10.1016/j.jad.2020.11.101](https://doi.org/10.1016/j.jad.2020.11.101)] [Medline: [33234283](https://pubmed.ncbi.nlm.nih.gov/33234283/)]
62. Ratcliffe GE, Enns MW, Belik SL, Sareen J. Chronic pain conditions and suicidal ideation and suicide attempts: an epidemiologic perspective. *Clin J Pain* 2008;24(3):204-210. [doi: [10.1097/AJP.0b013e31815ca2a3](https://doi.org/10.1097/AJP.0b013e31815ca2a3)] [Medline: [18287825](https://pubmed.ncbi.nlm.nih.gov/18287825/)]
63. Smith MT, Edwards RR, Robinson RC, Dworkin RH. Suicidal ideation, plans, and attempts in chronic pain patients: factors associated with increased risk. *Pain* 2004 Sep;111(1-2):201-208. [doi: [10.1016/j.pain.2004.06.016](https://doi.org/10.1016/j.pain.2004.06.016)] [Medline: [15327824](https://pubmed.ncbi.nlm.nih.gov/15327824/)]
64. Hanks RA, Rapport LJ, Vangel S. Caregiving appraisal after traumatic brain injury: the effects of functional status, coping style, social support and family functioning. *NeuroRehabilitation* 2007;22(1):43-52. [Medline: [17379947](https://pubmed.ncbi.nlm.nih.gov/17379947/)]
65. Rasic D, Weerasinghe S, Asbridge M, Langille DB. Longitudinal associations of cannabis and illicit drug use with depression, suicidal ideation and suicidal attempts among Nova Scotia high school students. *Drug Alcohol Depend* 2013 Apr 01;129(1-2):49-53. [doi: [10.1016/j.drugalcdep.2012.09.009](https://doi.org/10.1016/j.drugalcdep.2012.09.009)] [Medline: [23041136](https://pubmed.ncbi.nlm.nih.gov/23041136/)]
66. Wastler HM, Bryan AO, Bryan CJ. Suicide attempts among adults denying active suicidal ideation: an examination of the relationship between suicidal thought content and suicidal behavior. *J Clin Psychol* 2022 Jun;78(6):1103-1117. [doi: [10.1002/jclp.23268](https://doi.org/10.1002/jclp.23268)] [Medline: [34674388](https://pubmed.ncbi.nlm.nih.gov/34674388/)]
67. Schneider RA, Chen SY, Lungu A, Grasso JR. Treating suicidal ideation in the context of depression. *BMC Psychiatry* 2020 Oct 08;20(1):497 [FREE Full text] [doi: [10.1186/s12888-020-02894-5](https://doi.org/10.1186/s12888-020-02894-5)] [Medline: [33032569](https://pubmed.ncbi.nlm.nih.gov/33032569/)]
68. Zisook S, Trivedi MH, Warden D, Lebowitz B, Thase ME, Stewart JW, et al. Clinical correlates of the worsening or emergence of suicidal ideation during SSRI treatment of depression: an examination of citalopram in the STAR\*D study. *J Affect Disord* 2009 Sep;117(1-2):63-73. [doi: [10.1016/j.jad.2009.01.002](https://doi.org/10.1016/j.jad.2009.01.002)] [Medline: [19217668](https://pubmed.ncbi.nlm.nih.gov/19217668/)]
69. Øien-Ødegaard C, Hauge LJ, Reneflot A. Marital status, educational attainment, and suicide risk: a Norwegian register-based population study. *Popul Health Metr* 2021 Jul 12;19(1):33 [FREE Full text] [doi: [10.1186/s12963-021-00263-2](https://doi.org/10.1186/s12963-021-00263-2)] [Medline: [34247635](https://pubmed.ncbi.nlm.nih.gov/34247635/)]

70. Liu L, Capaldi CA, Dopko RL. Suicide ideation in Canada during the COVID-19 pandemic. *Health Promot Chronic Dis Prev Can* 2021 Nov 10;41(11):378-391 [FREE Full text] [doi: [10.24095/hpcdp.41.11.06](https://doi.org/10.24095/hpcdp.41.11.06)] [Medline: [34569774](https://pubmed.ncbi.nlm.nih.gov/34569774/)]
71. Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, et al. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry* 2008 Feb;192(2):98-105 [FREE Full text] [doi: [10.1192/bjp.bp.107.040113](https://doi.org/10.1192/bjp.bp.107.040113)] [Medline: [18245022](https://pubmed.ncbi.nlm.nih.gov/18245022/)]
72. Teicher MH, Glod CA, Cole JO. Antidepressant drugs and the emergence of suicidal tendencies. *Drug Saf* 1993 Mar;8(3):186-212. [doi: [10.2165/00002018-199308030-00002](https://doi.org/10.2165/00002018-199308030-00002)] [Medline: [8452661](https://pubmed.ncbi.nlm.nih.gov/8452661/)]
73. Hjelmeland H, Jaworski K, Knizek BL, Marsh I. Problematic advice from suicide prevention experts. *Ethical Hum Psychol Psychiatry* 2019 Mar 05;20(2):79-85 [FREE Full text] [doi: [10.1891/1559-4343.20.2.79](https://doi.org/10.1891/1559-4343.20.2.79)]
74. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, STAR\*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR\*D: implications for clinical practice. *Am J Psychiatry* 2006 Jan;163(1):28-40. [doi: [10.1176/appi.ajp.163.1.28](https://doi.org/10.1176/appi.ajp.163.1.28)] [Medline: [16390886](https://pubmed.ncbi.nlm.nih.gov/16390886/)]
75. Baertschi M, Costanza A, Canuto A, Weber K. The dimensionality of suicidal ideation and its clinical implications. *Int J Methods Psychiatr Res* 2019 Mar;28(1):e1755 [FREE Full text] [doi: [10.1002/mpr.1755](https://doi.org/10.1002/mpr.1755)] [Medline: [30426604](https://pubmed.ncbi.nlm.nih.gov/30426604/)]
76. Gibbons RD, Brown CH, Hur K, Davis J, Mann JJ. Suicidal thoughts and behavior with antidepressant treatment: reanalysis of the randomized placebo-controlled studies of fluoxetine and venlafaxine. *Arch Gen Psychiatry* 2012 Jun;69(6):580-587 [FREE Full text] [doi: [10.1001/archgenpsychiatry.2011.2048](https://doi.org/10.1001/archgenpsychiatry.2011.2048)] [Medline: [22309973](https://pubmed.ncbi.nlm.nih.gov/22309973/)]
77. Oquendo MA, Galfalvy H, Russo S, Ellis SP, Grunebaum MF, Burke A, et al. Prospective study of clinical predictors of suicidal acts after a major depressive episode in patients with major depressive disorder or bipolar disorder. *Am J Psychiatry* 2004 Aug;161(8):1433-1441. [doi: [10.1176/appi.ajp.161.8.1433](https://doi.org/10.1176/appi.ajp.161.8.1433)] [Medline: [15285970](https://pubmed.ncbi.nlm.nih.gov/15285970/)]
78. Omary A, Richardson U, Chambers T. Covariate health determinants of suicidal ideation, suicide plans, and suicide attempts among African American males with and without major depressive episode. *Arch Suicide Res* (forthcoming) 2021 Oct 20:1-15. [doi: [10.1080/13811118.2021.1989350](https://doi.org/10.1080/13811118.2021.1989350)] [Medline: [34669554](https://pubmed.ncbi.nlm.nih.gov/34669554/)]
79. Abraham A, Jithesh A, Doraiswamy S, Al-Khawaga N, Mamtani R, Cheema S. Telemental health use in the COVID-19 pandemic: a scoping review and evidence gap mapping. *Front Psychiatry* 2021 Nov 8;12:748069 [FREE Full text] [doi: [10.3389/fpsy.2021.748069](https://doi.org/10.3389/fpsy.2021.748069)] [Medline: [34819885](https://pubmed.ncbi.nlm.nih.gov/34819885/)]
80. 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](https://doi.org/10.1037/a0027924)] [Medline: [22662727](https://pubmed.ncbi.nlm.nih.gov/22662727/)]
81. Batastini AB, Paprzycki P, Jones AC, MacLean N. Are videoconferenced mental and behavioral health services just as good as in-person? A meta-analysis of a fast-growing practice. *Clin Psychol Rev* 2021 Feb;83:101944. [doi: [10.1016/j.cpr.2020.101944](https://doi.org/10.1016/j.cpr.2020.101944)] [Medline: [33227560](https://pubmed.ncbi.nlm.nih.gov/33227560/)]
82. Li H, Glesia A, Kent-Wilkinson A, Leidl D, Kleib M, Risling T. Transition of mental health service delivery to telepsychiatry in response to COVID-19: a literature review. *Psychiatr Q* 2022 Mar;93(1):181-197 [FREE Full text] [doi: [10.1007/s11126-021-09926-7](https://doi.org/10.1007/s11126-021-09926-7)] [Medline: [34101075](https://pubmed.ncbi.nlm.nih.gov/34101075/)]
83. Balestrieri M, Sisti D, Rocchi M, Rucci P, Simon G, Araya R, et al. Effectiveness of clinical decision support systems and telemedicine on outcomes of depression: a cluster randomized trial in general practice. *Fam Pract* 2020 Nov 28;37(6):731-737. [doi: [10.1093/fampra/cmz077](https://doi.org/10.1093/fampra/cmz077)] [Medline: [32766705](https://pubmed.ncbi.nlm.nih.gov/32766705/)]
84. Kurian BT, Trivedi MH, Grannemann BD, Claassen CA, Daly EJ, Sunderajan P. A computerized decision support system for depression in primary care. *Prim Care Companion J Clin Psychiatry* 2009;11(4):140-146 [FREE Full text] [doi: [10.4088/PCC.08m00687](https://doi.org/10.4088/PCC.08m00687)] [Medline: [19750065](https://pubmed.ncbi.nlm.nih.gov/19750065/)]
85. Trivedi MH, Rush AJ, Crismon ML, Kashner TM, Toprac MG, Carmody TJ, et al. Clinical results for patients with major depressive disorder in the Texas Medication Algorithm Project. *Arch Gen Psychiatry* 2004 Jul;61(7):669-680. [doi: [10.1001/archpsyc.61.7.669](https://doi.org/10.1001/archpsyc.61.7.669)] [Medline: [15237079](https://pubmed.ncbi.nlm.nih.gov/15237079/)]

## Abbreviations

**GAD:** general anxiety disorder

**GAD-7:** Generalized Anxiety Disorder-7

**OR:** odds ratio

**PHQ:** Patient Health Questionnaire

**SI:** suicidal ideation

**STAR\*D:** Sequenced Treatment Alternatives to Relieve Depression

*Edited by A Mavragani; submitted 04.03.22; peer-reviewed by A Costanza, E Chan, E Palmer; comments to author 07.08.22; revised version received 16.08.22; accepted 06.09.22; published 30.09.22.*

*Please cite as:*

*O'Callaghan E, Mahrer N, Belanger HG, Sullivan S, Lee C, Gupta CT, Winsberg M*

*Telehealth-Supported Decision-making Psychiatric Care for Suicidal Ideation: Longitudinal Observational Study*

*JMIR Form Res 2022;6(9):e37746*

*URL: <https://formative.jmir.org/2022/9/e37746>*

*doi: [10.2196/37746](https://doi.org/10.2196/37746)*

*PMID: [36178727](https://pubmed.ncbi.nlm.nih.gov/36178727/)*

©Erin O'Callaghan, Nicole Mahrer, Heather G Belanger, Scott Sullivan, Christine Lee, Carina T Gupta, Mirène Winsberg. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 30.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Attention-Based Models for Classifying Small Data Sets Using Community-Engaged Research Protocols: Classification System Development and Validation Pilot Study

Brian J Ferrell<sup>1\*</sup>, MDA; Sarah E Raskin<sup>2\*</sup>, MPH, PhD; Emily B Zimmerman<sup>3\*</sup>, MS, MPH, PhD; David H Timberline<sup>1\*</sup>, MA; Bridget T McInnes<sup>4\*</sup>, PhD; Alex H Krist<sup>5\*</sup>, MPH, MD

<sup>1</sup>Center for Community Engagement and Impact, Virginia Commonwealth University, Richmond, VA, United States

<sup>2</sup>L Douglas Wilder School of Government and Public Affairs, Virginia Commonwealth University, Richmond, VA, United States

<sup>3</sup>Center on Society and Health, Virginia Commonwealth University, Richmond, VA, United States

<sup>4</sup>Computer Science Department, Virginia Commonwealth University, Richmond, VA, United States

<sup>5</sup>Department of Family Medicine and Population Health, Virginia Commonwealth University, Richmond, VA, United States

\* all authors contributed equally

**Corresponding Author:**

Brian J Ferrell, MDA

Center for Community Engagement and Impact

Virginia Commonwealth University

907 Floyd Ave

Richmond, VA, 23284

United States

Phone: 1 8043505426

Email: [ferrellbj@vcu.edu](mailto:ferrellbj@vcu.edu)

## Abstract

**Background:** Community-engaged research (CEnR) is a research approach in which scholars partner with community organizations or individuals with whom they share an interest in the study topic, typically with the goal of supporting that community's well-being. CEnR is well-established in numerous disciplines including the clinical and social sciences. However, universities experience challenges reporting comprehensive CEnR metrics, limiting the development of appropriate CEnR infrastructure and the advancement of relationships with communities, funders, and stakeholders.

**Objective:** We propose a novel approach to identifying and categorizing community-engaged studies by applying attention-based deep learning models to human participants protocols that have been submitted to the university's institutional review board (IRB).

**Methods:** We manually classified a sample of 280 protocols submitted to the IRB using a 3- and 6-level CEnR heuristic. We then trained an attention-based bidirectional long short-term memory unit (Bi-LSTM) on the classified protocols and compared it to transformer models such as Bidirectional Encoder Representations From Transformers (BERT), Bio + Clinical BERT, and Cross-lingual Language Model-Robustly Optimized BERT Pre-training Approach (XLM-RoBERTa). We applied the best-performing models to the full sample of unlabeled IRB protocols submitted in the years 2013-2019 (n>6000).

**Results:** Although transfer learning is superior, receiving a 0.9952 evaluation F1 score for all transformer models implemented compared to the attention-based Bi-LSTM (between 48%-80%), there were key issues with overfitting. This finding is consistent across several methodological adjustments: an augmented data set with and without cross-validation, an unaugmented data set with and without cross-validation, a 6-class CEnR spectrum, and a 3-class one.

**Conclusions:** Transfer learning is a more viable method than the attention-based bidirectional-LSTM for differentiating small data sets characterized by the idiosyncrasies and variability of CEnR descriptions used by principal investigators in research protocols. Despite these issues involving overfitting, BERT and the other transformer models remarkably showed an understanding of our data unlike the attention-based Bi-LSTM model, promising a more realistic path toward solving this real-world application.

(JMIR Form Res 2022;6(9):e32460) doi:[10.2196/32460](https://doi.org/10.2196/32460)



## KEYWORDS

data augmentation; BERT; transformer-based models; text classification; community engagement; prototype; IRB research; community-engaged research; participatory research; deep learning

## Introduction

Transfer learning is widely used when comparing traditional machine learning and deep learning models [1]. It is likely that transformer models like Bidirectional Encoder Representations From Transformers (BERT) [2], a neural network-based technique for natural language processing (NLP) pretraining, will always play a substantial part in how we model language [3]. Researchers attempt to make use of these language models and fine-tune them to their classification tasks using various data sets. Superior results have been found with large data sets [4], small data sets [5,6], short text lengths [7], longer text lengths [8], and even data sets of different languages [1]. These studies, and the work reported here, demonstrate that better results can be achieved without substantial amounts of computing power and data.

Community-engaged research (CEnR) is a research approach in which investigators from conventional research institutions, such as universities, partner with community members or organizations with whom they share an interest, typically with the goal of advancing that community's well-being [9]. Defined by its research philosophy and the relationship between research partners rather than methodology, CEnR is now an established scholarly tradition in numerous disciplines including health sciences, the social sciences, social work, urban planning, education, and the arts. Teams using CEnR have implemented research projects addressing a wide range of stakeholder concerns; collaborated with partners across the research process [10-13], from problem identification to scaling evidence-based interventions [14]; transformed service learning with new curricula and pedagogies that reflect students' interests and learning styles [15]; and transformed natural, built, and artistic environments to better reflect the values and interests of communities [16].

CEnR's flexibility and breadth has been productive, resulting in dedicated journals, conferences, courses, funding mechanisms, evaluation metrics, and theories of classification along continua of activities and structures of governance. Yet identifying, describing, measuring, and reporting on CEnR studies in the aggregate has been a challenge for universities and other institutions (eg, disciplinary associations [17]), in particular, reporting valid and reliable metrics to funders and stakeholders [17], and developing and maintaining appropriate internal CEnR infrastructure. Dependence on conventional review mechanisms such as scholarly databases to provide data on CEnR productivity may be limited by diversity in disciplines, methods, and dissemination approaches; impacts that are primarily shared outside of traditional scholarly mechanisms such as peer-reviewed journals; and inaccurate selection of CEnR as a keyword. The limited federal and foundation support available for CEnR obviates searches of funding databases. Moreover, established mechanisms for identifying and tracking CEnR may privilege recognition of CEnR collaborations that proceed along a unidirectional pathway in which relationships

between professional researchers and community members demonstrate a deepening collaboration over time, resulting in grants and peer-reviewed publications. Such an emphasis both belies the reality of inequities in the distribution of resources needed to sustain such collaborations, for example, between disciplines, between research-productive and teaching institutions, and between established and junior faculty.

Virginia Commonwealth University (VCU) is an R01 institution designated by the Carnegie Foundation as "Community Engaged" with "Highest Research Activity." In 2013, VCU began flagging CEnR studies using three custom fields [18] in the university's online human participants protocol submission form, as part of an award from the National Center for Advancing Translational Sciences.

- Is there at least one community partner involved in the proposed study? (Yes/no answer)
- If yes, who is the community partner?
  - Name of organization
  - Zip code or country of the organization
- Which of the three statements below best describes the role of the community partner in the study?
  - Community partners only provide access to study participants or project sites. They are not involved with study design, participant recruitment, data collection, or data analysis.
  - Community partners do not make decisions about the study design or conduct but provide guidance to the researcher about the study design, participant recruitment, data collection, or data analysis.
  - Community partners make decisions with the researchers about the study's research activities or help conduct those activities (ie, study design, participant recruitment, data collection, or data analysis) [19].

Technical impediments to entering data into these custom fields were identified in 2018. This quality concern initiated a broader discussion among stakeholders across VCU about other possible limitations in the system of documentation, for example, inconsistent interpretation of these fields by principal investigators or study administrators submitting protocols. This discussion led to the exploratory study described here. The overall aim of this study was to develop a methodology to automatically detect CEnR studies among protocols submitted in the university's online institutional review board (IRB) system, which contains data on all research with human participants [20]. This study provided the opportunity to test and build on the three custom fields added to the IRB protocol. The subaims are as follows: develop a system of classification to adapt the conventional theorization of CEnR across a spectrum of collaboration to the practical reality of studies conducted at an R01 university, determine if one or more deep learning models could automate the identification of CEnR studies trained by a subset of hand-labeled IRB protocols, and

identify the best-performing algorithms and apply them to a retrospective 5-year data set of unlabeled research protocols ( $n > 6000$ ) that were not incorporated in the training of the algorithm.

## Methods

---

### Data

#### *Data Collection*

The first stage of this process was to pull research protocols from the IRB's database ( $n > 20,000$ ). We then cleaned and deduplicated the records (1 study per protocol, "exempt,"

"expedited," "full," and "started/submitted" protocols were included, but "not yet reviewed" studies were left out), leaving us with 6000 research studies, from which a sample ( $n = 280$ ) was randomly selected, reviewed, and manually labeled as one of the six classes (described in the Data Annotation section). Our criteria for selecting this sample set were based on a research study's likelihood of being CEnR or not. [Textbox 1](#) shows the chosen columns and a snippet of what the data looks like. Examples of the terminology we used for finding potential CEnR research are as follows: community-engaged, community-based participatory research, (community) action research, participatory action research, community advisory group, community steering, etc.

**Textbox 1.** The institutional review board protocol fields used to classify protocols with brief example sentences. These fields were concatenated into one column during training.

<p><b>Study title</b></p> <ul style="list-style-type: none"> <li>• “Exploring dental service underutilization among...”</li> <li>• “Regional Scan and Strategies for Community Eng...”</li> <li>• “Reflections on 5 years of community-based part...”</li> </ul> <p><b>Informed personnel</b></p> <ul style="list-style-type: none"> <li>• “The research team is in routine contact among...”</li> <li>• “The team has three weekly meetings to inform t...”</li> <li>• “We are a research team that collaborates on a...”</li> </ul> <p><b>Scientific benefit</b></p> <ul style="list-style-type: none"> <li>• “This research is intended to identify, describ...”</li> <li>• “This study is meant to inform community leader...”</li> <li>• “This study will address gaps in scientific know...”</li> </ul> <p><b>Aims and goals</b></p> <ul style="list-style-type: none"> <li>• “The overall aim of this mixed methods study is...”</li> <li>• “Based on the results of the literature review,...”</li> <li>• “The goal is to describe and publish the effect...”</li> </ul> <p><b>Identify participants</b></p> <ul style="list-style-type: none"> <li>• “ALL PARTICIPANTS Community Partner has experience admin...”</li> <li>• “We will first scan regional organizational to...”</li> <li>• “We already have contact and working relationsh...”</li> </ul> <p><b>Background</b></p> <ul style="list-style-type: none"> <li>• “Unmet dental needs are significant public heal...”</li> <li>• “This project is part of a larger Richmond init...”</li> <li>• “The field of CBPR still suffers from gap in e...”</li> </ul> <p><b>Hypothesis</b></p> <ul style="list-style-type: none"> <li>• “As a mixed-methods study, this research uses a...”</li> <li>• “This project is to complete a literature revie...”</li> <li>• “We are trying to document the direct and indir...”</li> </ul> <p><b>Study design</b></p> <ul style="list-style-type: none"> <li>• “STUDY DESIGN This mixed methods study is a cros...”</li> <li>• “Regional Scan For the regional scan, the projec...”</li> <li>• “We will talk to selected community partners an...”</li> </ul>
--

### Data Annotation

We uploaded the newly extracted sample data set into Google Sheets to facilitate a collaborative process of manually reviewing and labeling the protocols for use in training the algorithm. The team of three reviewers (two per research study) reviewed the available data for each protocol and labeled it “yes” (CEnR) or “no” (not CEnR) and assigned a class corresponding to the CEnR level (0-6). Protocols that did not receive the same

designation by both reviewers were discussed and resolved in weekly meetings.

### CEnR Levels

After a preliminary review of the protocols, the reviewers inductively developed a coding system to reflect the types of relationships described in the protocols. [Textbox 2](#) shows a breakdown of CEnR levels that were used by reviewers.

**Textbox 2.** CEnR levels that were used to manually classify the training data.

**No community-engaged research (CEnR; 0)**

Research without a partnership or community engagement

**Non-CEnR partnership (1)**

There is reference to a partnership, but the relationship is uncategorizable (eg, not adequately described) or not a traditional community-engaged partnership (eg, contractual relationships).

**Instrumental partnership (2)**

The community partner primarily facilitates access to the “inputs” needed to conduct the study (eg, posting recruitment flyers, providing participant contact information, extracting data, or providing study sites for observation).

**Academic-led partnership (3)**

Minimal yet important interaction between the research team and the community partner, which is often essential to project success (eg, academic partners take the lead on study design and research activities, with community partner involvement at particular points, such as troubleshooting recruitment or facilitating community meetings)

**Cooperative partnership (4)**

Shared investment and mutual consideration between the research team and the community partner, without shared decision-making (eg, community advisory boards that provided input on study design methodology, reviewed data collection instruments, interpreted findings, or informed dissemination plans)

**Reciprocal partnership (5)**

Community partners and research teams share decision-making power and governance (eg, community-based participatory research, team science, or steering committees with decision-making power).

### Data Cleaning

After reviewing and classifying the protocols, we checked again for duplications, did manual spell-checking, and trimmed white space and any irrelevant symbols. Final data cleaning was completed with Python using the NLTK package (stop words, lemmatization, lowercase, removing punctuation, splitting contractions, and other RegEx operations).

### Data Augmentation

We tested whether data augmentation techniques [21] (replacing and inserting words [22]) using the nlpaug library [23] to synthetically increase the amount of training data using DistilBERT [24] would improve the performance. Table 1 shows the number of samples before and after augmentation.

**Table 1.** Number of examples per class before and after data augmentation (second data set).

Class	Samples (before), n	Samples (after), n
0	82	1931
1	40	1427
2	11	1413
3	101	1564
4	32	1431
5	13	1404

### Data Sets

We used three data sets: (1) the original sample of 280 hand-classified protocols, (2) an augmented data set of the 280 protocol expanded to 9170 samples using DistilBERT, and (3) versions of the first two data sets with 6 classes merged into 3. We tested the data set with fewer categories of CEnR to explore whether using broader categories would improve generalization of the models and prediction score. For data sets containing three classes, we collapsed 1s and 2s (=1); collapsed 3s, 4s, and 5s (=2); and kept the class 0 as is.

### Models

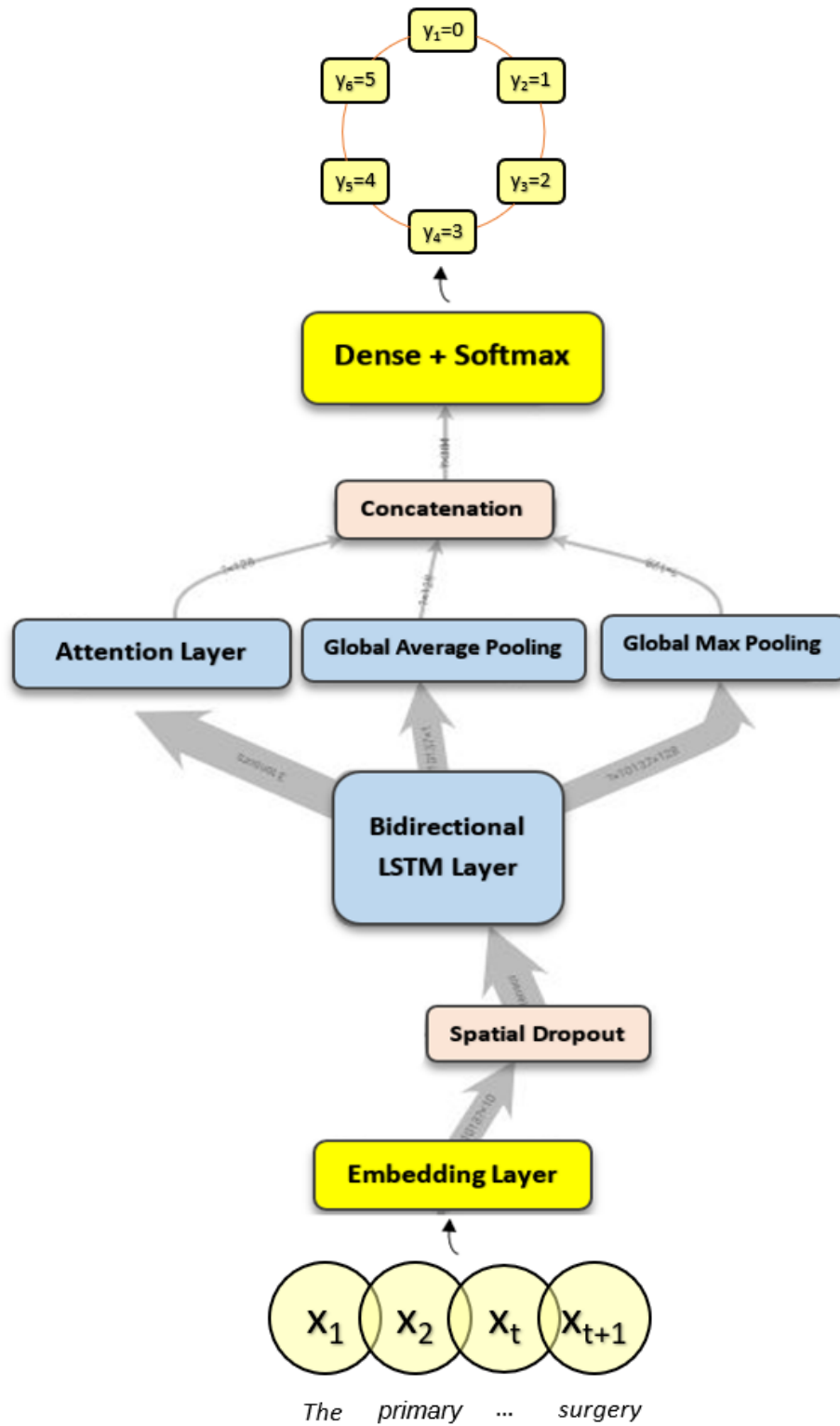
We explored four models to classify the data into the CEnR classes: bidirectional long short-term memory unit (Bi-LSTM),

BERT, Bio + Clinical BERT, and Cross-lingual Language Model–Robustly Optimized BERT Pre-training Approach (XLM-RoBERTa) transformer models. We present model architectures and hyperparameters in this section.

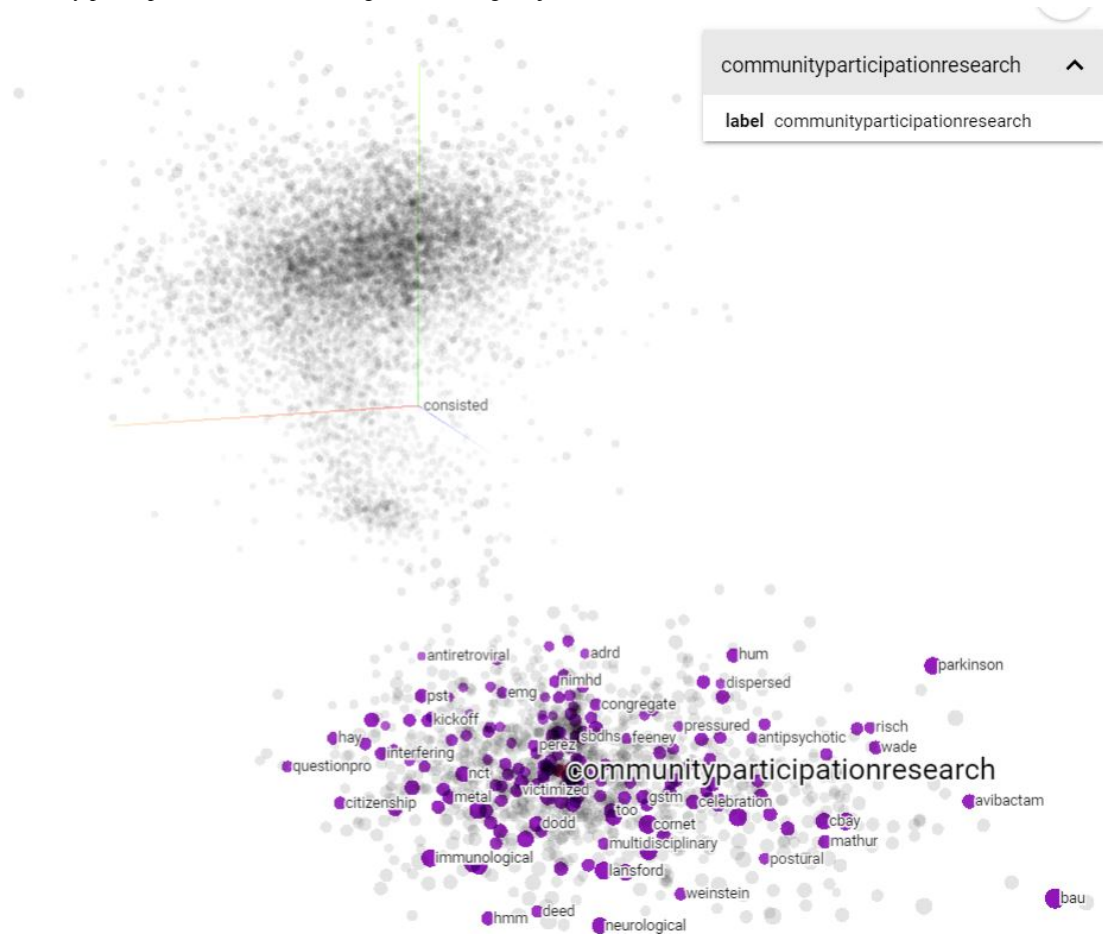
### Bi-LSTM Attention Model

Figure 1 illustrates the first model: a Bi-LSTM [25-27] with a basic custom attention layer [28,29] that was concatenated with a GlobalMaxPooling and GlobalAveragePooling layer. The embeddings used were the 100-dim Global Vectors for Word Representation (GloVe) embeddings file containing 400,000 words computed on a 2014 dump of English Wikipedia [30]. GloVe is an unsupervised learning algorithm for retrieving vector representations of words that can be plotted in a geometric space [31], as seen in Figure 2.

**Figure 1.** Attention-based bidirectional LSTM model architecture. LSTM: long short-term memory unit.



**Figure 2.** Searched “community participation research” in Google Embedding Projector.

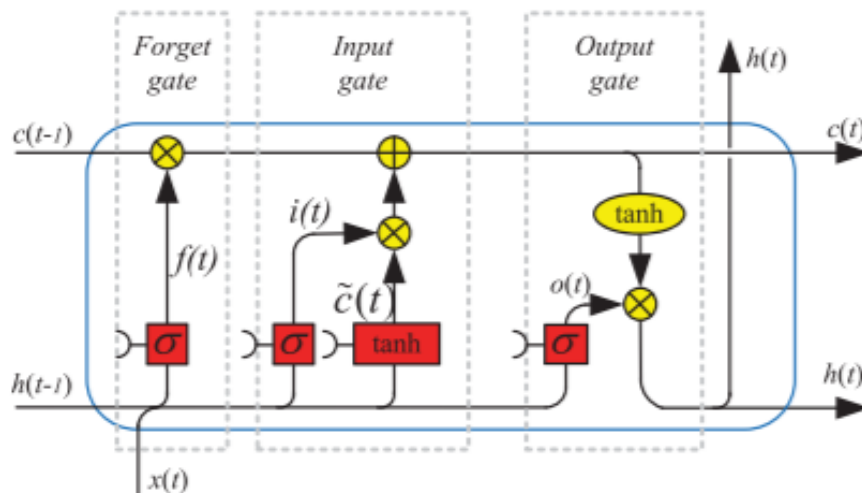


The embedding layer captures the similarities between the words to best optimize for our inputs, and the Bi-LSTM runs through the data from the beginning of a sentence to the end and vice versa. This is done through its four [32] components as seen in Figure 3: cell state ( $C_t$ ), forget gate ( $f_t$ ), input gate ( $i_t$  and  $\square$ ), and output gate ( $O_t$  and  $h_t$ ). These control the flow of sequential information, regulating what is important and what is not from those embeddings. The attention layer (which adds a weight of importance [33] to those Bi-LSTM outputs), the max pooling layer (which finds the most important features from the Bi-LSTM outputs), and the average pooling layer (which weighs all outputs from the Bi-LSTM as important) become fused together into one matrix to give the neural network more features to base predictions on. Finally, a dense layer with the softmax

function is the flow of calculations made to give us a final output of a  $Y=[0,1,2,3,4,5]$  classification.

Stratified 7-fold cross-validation, Synthetic Minority Oversampling Technique (SMOTE) [34], and F1 macro optimization [35] were also used. Stratified K-fold cross-validation ensures the distribution of classes remains the same in every fold. SMOTE is a way to create fake data for the minority classes using examples that are similar (k-nearest neighbors). This technique was used within folds of cross-validation during training, not before. F1 macro optimization ensures that the F1 score is optimized during training, not accuracy. F1 macro refers to the average of the class's F1 scores; this technique increased our evaluation F1 score by 7%.

**Figure 3.** Architecture of long short-term memory unit.



### Transformer Models

Transfer learning takes large and powerfully built language models that are pretrained on large corpuses of unlabeled data to later be fine-tuned and repurposed for a second related task, which can be beneficial for small data sets. A main aspect of this study was to see if the use of transfer learning improved the predictive performance for our text classification task. We used BERT-base-uncased [2], Bio + Clinical BERT [36], and XLM-RoBERTa [37] models, and tried different learning rates, batch sizes, and epochs for all three separately (around 30-50 different models trained per transformer). The Results section shows the best-tuned model for each transformer.

### Bidirectional Encoder Representations From Transformers

Our first approach to transfer learning was fine-tuning the pretrained BERT model for our text classification problem. BERT was introduced by Devlin et al [2]. It was pretrained on BookCorpus (800 million words) and Wikipedia (2500 million words). The model's architecture ensures its advantage in NLP tasks because it learns the contextual meanings of words and how each word is being used in a sequence due to its 12 attention heads and 110 million total parameters. GloVe embeddings do not consider the context of how a word is used and do not capture the different semantics that words can have (eg, a bat can be an animal or baseball equipment); thus the word "community" or "partner" can be used differently across different research studies. BERT, however, would capture those differences. Additionally, BERT can achieve state-of-the-art results on various tasks for large and small data sets, and it does not need to be trained for more than 2 to 4 epochs.

### BIO + Clinical BERT

The second approach to transfer learning is fine-tuning with Bio + Clinical BERT [36]. As mentioned previously, BERT is pretrained on BookCorpus and Wikipedia, and in general can model language well for any NLP task; however, Alsentzer et al [36] examined ways to improve the general language model in BERT using BERT models geared for clinical text and discharge summaries. They demonstrated that performance is improved with domain-specific pretrainings, which is distinct

from general language. The authors used data from the MIMIC-III database in two ways, clinical BERT (contains all note types) and discharge summary BERT (only contains discharge summaries), to further downstream tasks with clinical data that can be used for more specific classification problems. They then trained two BERT models on the clinical text, where one is initialized from the BERT-base model and the other was initialized from BioBERT (the model we chose).

### Cross-lingual Language Model–Robustly Optimized BERT Pre-training Approach

Our third approach to transfer learning was an interesting model to fine-tune, mainly because this type of transformer model was not created for our kind of task; however, it still performed well. It was introduced by Conneau et al [37] in 2019 and updated in 2020. This model closely resembles the RoBERTa architecture [38], except it is a cross-lingual model pretrained on 100 different languages. This type of model is made for cross-lingual transfer learning tasks trained on more than 2 terabytes of the CommonCrawl corpora.

### Other Models

Other models were used for this study, such as convolutional neural networks (CNNs), deep neural networks (DNNs), CNN + LSTM, CNN + Bi-LSTM, CNN + Bi-LSTM with attention, CNN + LSTM with attention, CNN + gated recurrent unit (GRU), CNN + Bi-GRU, CNN + Bi-GRU with attention, and CNN + GRU with attention; however, they did not perform as well as the Bi-LSTM + attention (ranging from a 0.30-0.40 evaluation F1 scores); therefore, we did not include their results in this paper.

### Experimental Details

#### Bi-LSTM Attention Model

In this model, we used the Keras libraries for training, tokenizing, and padding the sequences of text. The Bi-LSTM model was trained for 40 epochs, had a learning rate of 0.001, batch size of 64, and was trained for 12 hours; additionally, we used the Adam optimizer and sparse categorical cross entropy for our loss. The max sequence length after cleaning the data

was 10,137. The model was trained as a CuDNNLSTM, which is a faster implementation of the LSTM backed up by CuDNN, which can only be run on a GPU.

### Transformer Models

We used the SimpleTransformers library created by Rajapakse [39], which can train and evaluate transformer models (derived from the HuggingFace web site) with few lines of code. The hyperparameters for each transformer model can be seen from a web site called Weights and Biases that organizes and captures all the necessary data during training [40,41]. Since the text field lengths in our sample were longer than the limits for BERT and other transformer models, we used a sliding window technique. Here, any sequence from the data that exceeds the maximum sequence length will be split into several subsets, each pertaining to the length of the max sequence length value. Using this technique, each subset from the sliding window has overlapping values, also referred to as the stride (stride 0.8) resulting in about a 20% overlap between the windows. This process lengthens training time but is preferable to truncating data during training. All models were trained using Google Colab Pro and had weights corresponding to a class so that it was equally balanced during the training [42].

### Evaluation Metrics

The models trained were evaluated using the F1 score macro, which takes a balanced measure of precision and recall, and then the average of the F1 scores.

## Results

Table 2 shows the holdout F1 scores for each of our models on our original and augmented data sets with and without cross-validation. The evaluation F1 scores (not shown in the table) for the Bi-LSTM averaged 63.25%. From the order of Table 2, it was 65% (with cross-validation, augmented) and 48% (without cross-validation, augmented) for 6 classes, and 80% (with cross-validation, augmented) and 60% (without cross-validation, augmented) for 3 classes, whereas the transformer model's evaluation F1 scores were all over 99%. We used Bio + Clinical BERT because domain-specific pretrainings have been shown to improve performance [34], and because our data set contains clinical research data, we thought it was relevant to compare its results. XLM-RoBERTa proved to do well and had an overall great understanding of the data, so it was included in this experiment as well. The holdout data set comprises 30 samples, which is almost too small to give an accurate account of how the models do, so our team will be working on labeling additional data. It is also a bit deceptive with the results shown because the classifications for the Bi-LSTM attention model were way off, whereas when the transformer models misclassified a research study, it was off by 1 or 2 classes. A lot of the results are not shown in the table. This is because it was not worth training the original data set without cross-validation due to the data set's size, which would also make the evaluation data set different, and there was no training for Bio + Clinical BERT and XLM-RoBERTa for augmented data sets using cross-validation due to computational limitations.

**Table 2.** Results of the various models over the original and augmented data sets.

Model	Data	6 classes, F1 scores		3 classes, F1 scores	
		With CV <sup>a</sup>	Without CV	With CV	Without CV
Bi-LSTM <sup>b</sup> w/ attention	Original	0.2000	N/A <sup>c</sup>	0.3000	N/A
Bi-LSTM w/ attention	Augmented	0.2667	0.3000	0.4000	0.2667
BERT <sup>d</sup> -base uncased	Original	0.2333	N/A	0.5000	N/A
BERT-base uncased	Augmented	0.3333	0.4000	0.4667	0.5333
Bio + Clinical BERT	Original	0.3000	N/A	0.4667	N/A
Bio + Clinical BERT	Augmented	N/A	0.4000	N/A	0.4333
XLM-RoBERTa <sup>e</sup>	Original	0.3667	N/A	0.4667	N/A
XLM-RoBERTa	Augmented	N/A	0.4000	N/A	0.4667

<sup>a</sup>CV: cross-validation.

<sup>b</sup>Bi-LSTM: bidirectional long short-term memory unit.

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

<sup>d</sup>BERT: Bidirectional Encoder Representations From Transformers.

<sup>e</sup>XLM-ROBERTa: Cross-lingual Language Model–Robustly Optimized BERT Pre-training Approach.

## Discussion

### Principal Findings

The transformer models performed significantly better than the Bi-LSTM with attention. They were nearly perfect for their evaluation scores (all hitting 0.995) across all the data sets used

(they overfit on the holdout data sets due to the same learning rate being used for each layer). Additionally, all models showed slight improvements when the number of classes fit a 3-class spectrum as opposed to a 6-class spectrum. It was hard to tell if the augmented data sets gave an advantage to the models; therefore, there is a need to research other techniques for that.



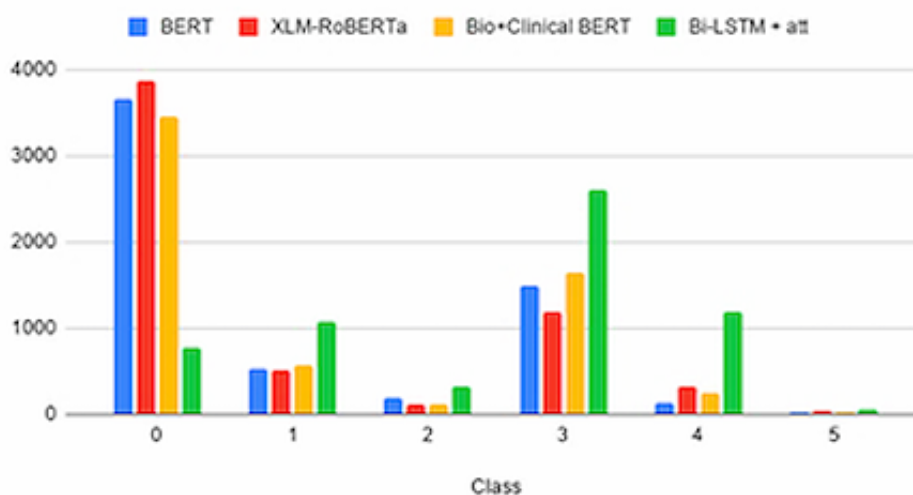
Cross-validation for the Bi-LSTM significantly improved its results for the evaluation scores but that did not carry over into the holdout data sets. The best-performing models for the 6-class spectrum was a 3-way tie between the transformer models that did not use cross-validation. Cross-validation was not needed when using the augmented data sets in terms of their holdout set scores. Although the BERT model trained on the augmented data set without using cross-validation had superior performance (0.533 holdout F1 score), the second best-performing model (BERT trained on the original data set with cross-validation) with less data trained much faster, and the results differed only fractionally compared to the best-performing one. We believe that data augmentation has great potential (considering it gives more data), and it may confer advantages during a model's training, but we feel it is better to go without it until more strategies are investigated. The strategies used were a faster way of synthetically creating more data, which does not necessarily mean it was the best way.

The Bi-LSTM attention model did not delineate between the classes nearly as well as BERT and the other transformer models, which has given our team a proof of concept, something to work with and improve on moving forward, whether that be more data or more computational power. Additionally, since there were only minor differences within the research study's augmentations (simple replacing and inserting of contextual similar words), BERT and the other transformers were able to pick up on those patterns almost perfectly compared to the Bi-LSTM model.

This study demonstrates that transfer learning performed better for classifying levels of CEnR. However, the results for the holdout sets were still relatively low (highest was 0.533), which we hope to improve with an increased data set size. We were impressed by the efficiency of BERT and other transformer models. While it took months of testing to identify the approach for using the Bi-LSTM with attention, and even more time to tune the hyperparameters, in a single day, BERT was able to achieve performances like the results shown in Table 2, with a significant decrease in training time. Considering those advantages, transfer learning appears to come out on top when it comes to hyperparameter selection.

The transformer model's final predictions versus the Bi-LSTM's final predictions on the remaining unlabeled data set are shown in Figure 4. The figure shows that predictions with the highest levels of engagement (4s and 5s) were lower from the transfer learning models, indicating a better understanding of our data in the real world, where 4s and 5s are infrequent in the data set and most protocols are zeros. This is the case because the IRB database represents all types of research, of which CEnR is a relatively small fraction. Bio + Clinical BERT and XLM-RoBERTa had results that were like BERT, although BERT was arguably more realistic. Of the transformer models, they agree on almost 4000 research studies' predictions; however, the attention-based model is only in agreement with all of them 850 (of the 6000) times.

**Figure 4.** Model predictions on 6000 research studies. att: attention; BERT: Bidirectional Encoder Representations From Transformers; Bi-LSTM: bidirectional long short-term memory unit; XLM-ROBERTa: Cross-lingual Language Model–Robustly Optimized BERT Pre-training Approach.



## Limitations

Researchers had the option of attaching detailed protocols as a PDF file instead of filling out the database fields. We were not able to retrieve PDF data for this study, reducing the total number of studies, which limited what data we could label. In addition, we observed that the transformer models predicted larger classes compared to smaller classes (eg, levels two, four, and five). Nevertheless, they still made reasonable predictions, which is exciting to see because it means we can improve from this issue moving forward by labeling more data or sticking to

the 3-class spectrum. We were also limited in our ability to compute very large models when using Google Colab Pro, which has certain computing limitations. Another time-consuming step was reviewing and labeling the data. The transformer models were derived from a library in which the overall structure is in its basic form; therefore, more adjustments can be made on their architectures [4,8].

## Conclusions

In conclusion, we compared widely used techniques in classification tasks: transfer learning using BERT, Bio + Clinical

BERT, XLM-RoBERTa, and a Bi-LSTM attention model. We found that transfer learning performed best for our purposes and was quick and easy to implement. Additional work is needed to apply the model in a system. In terms of process, we found that augmenting the data set has the potential to improve the results, cross-validation was not as helpful for the transformer models when using a less general classification spectrum, hyperparameter tuning with transformer models was less stressful and time-consuming, transformer models can handle small data sets well, and condensing the 6 classes into 3 was a less rigid spectrum for models to differentiate and provided superior results.

Additional improvements can be made, such as correcting a sample from the final prediction's data set by using the same search word criteria as before (Data Collection section) or by taking a random sample to increase our training data. We could also use different augmentation techniques, as there are other ways this could have been implemented. Future work includes

fine-tuning strategies and hyperparameter optimization such as discriminative learning rates, slanted triangular learning rates, and freezing layers. BERT is the best model from this study mainly because of its holdout score for the 3-class spectrum, and its training time is much faster than the other two transformer models; however, moving forward, all three transformer models will continue to be used in improving this experiment, as each is unique in their understanding of the data.

Identifying CEnR and classifying levels of engagement allow us to understand the types of research taking place across the university. These data can help organizations better serve their stakeholders and to plan for the infrastructure needed to support community engagement. Additionally, tracking these metrics can help institutions report to funders and stakeholders on their engagement activities. The innovative aspect of this methodological study is creating an automated system to categorize research using administrative data. This study describes how transformer models can automate this process.

---

## Acknowledgments

Our work has been supported by the National Institutes of Health (grant CTSA UL1TR002649, National Center for Advancing Translational Sciences).

---

## Authors' Contributions

BJF, SER, and EBZ conceptualized the study and designed the methodology. BJF and DHT used the software. BJF performed the formal analysis. BJF, SER, EBZ, and DHT performed the investigation. BJF and DHT curated the data. BJF wrote the original draft. BJF created the visualization of the data. BJF, BTM, and AHK supervised the study. BJF, SER, and EBZ were project administrators for the study. BJF and DHT provided the resources for the study. BJF and BTM validated the study. SER, EBZ, BTM, and AHK reviewed and edited the paper.

---

## Conflicts of Interest

None declared.

---

## References

1. González-Carvajal S, Garrido-Merchán EC. Comparing BERT against traditional machine learning text classification. arXiv Preprint posted online on May 26, 2020. [[FREE Full text](#)]
2. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online on October 11, 2018. [[FREE Full text](#)]
3. Wang C, Li M, Smola AJ. Language models with transformers. arXiv Preprint posted online on April 20, 2019. [[FREE Full text](#)]
4. Colón-Ruiz C, Segura-Bedmar I. Comparing deep learning architectures for sentiment analysis on drug reviews. *J Biomed Inform* 2020 Oct;110:103539 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2020.103539](https://doi.org/10.1016/j.jbi.2020.103539)] [Medline: [32818665](https://pubmed.ncbi.nlm.nih.gov/32818665/)]
5. Balagopalan A, Eyre B, Rudzicz F, Novikova J. To BERT or not to BERT: comparing speech and language-based approaches for Alzheimer's disease detection. arXiv Preprint posted online on July 26, 2020. [[FREE Full text](#)]
6. Usherwood P, Smit S. Low-shot classification: a comparison of classical and deep transfer machine learning approaches. arXiv Preprint posted online on July 17, 2019. [[FREE Full text](#)]
7. Ranti D, Hanss K, Zhao S, Arvind V, Titano J, Costa A, et al. The utility of general domain transfer learning for medical language tasks. arXiv Preprint posted online on February 16, 2020. [[FREE Full text](#)]
8. Sun C, Qiu X, Xu Y, Huang X. How to fine-tune BERT for text classification? In: Sun M, Huang X, Ji H, Liu Z, Liu Y, editors. Chinese Computational Linguistics 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings. Cham: Springer; 2019:194-206.
9. Principles of community engagement - second edition. Agency for Toxic Substances and Disease Registry, Centers for Disease Control and Prevention. 1997. URL: <https://www.atsdr.cdc.gov/communityengagement/> [accessed 2022-08-28]
10. Bers MU, González-González C, Armas-Torres MB. Coding as a playground: promoting positive learning experiences in childhood classrooms. *Comput Education* 2019 Sep;138:130-145. [doi: [10.1016/j.compedu.2019.04.013](https://doi.org/10.1016/j.compedu.2019.04.013)]

11. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. *JMIR Mhealth Uhealth* 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: [10.2196/14512](https://doi.org/10.2196/14512)] [Medline: [31934874](https://pubmed.ncbi.nlm.nih.gov/31934874/)]
12. Mosavel M, Ferrell D, LaRose JG, Lu J, Winship J. Conducting a community "Street Survey" to inform an obesity intervention: the WE Project. *Fam Community Health* 2021;44(3):117-125 [FREE Full text] [doi: [10.1097/FCH.0000000000000271](https://doi.org/10.1097/FCH.0000000000000271)] [Medline: [32842000](https://pubmed.ncbi.nlm.nih.gov/32842000/)]
13. Mosavel M, Gough MZ, Ferrell D. Using asset mapping to engage youth in community-based participatory research: The WE Project. *Prog Community Health Partnersh* 2018;12(2):223-236. [doi: [10.1353/cpr.2018.0042](https://doi.org/10.1353/cpr.2018.0042)] [Medline: [30270232](https://pubmed.ncbi.nlm.nih.gov/30270232/)]
14. Zimmerman EB. *Researching Health Together: Engaging Patients and Stakeholders, From Topic Identification to Policy Change*. Newbury Park, CA: SAGE Publications; 2020.
15. Richards-Schuster K. In: Cammarota J, Fine M, editors. *Revolutionizing Education: Youth Participatory Action Research in Motion*. Abingdon, Oxfordshire: Routledge; 2008.
16. Vanzandt S, Materson JH, Newman GD, Meyer MA. *Engaged Research for Community Resilience to Climate Change*. Amsterdam: Elsevier; 2020.
17. Eder MM, Evans E, Funes M, Hong H, Reuter K, Ahmed S, et al. Defining and measuring community engagement and community-engaged research: clinical and translational science institutional practices. *Prog Community Health Partnersh* 2018;12(2):145-156 [FREE Full text] [doi: [10.1353/cpr.2018.0034](https://doi.org/10.1353/cpr.2018.0034)] [Medline: [30270224](https://pubmed.ncbi.nlm.nih.gov/30270224/)]
18. Holton V. IRB-approved community-engaged research at VCU: analysis of baseline data and overview of new data system. *Community-Engaged Res Resources* 2013:1.
19. Holton VL, Jettner JF, Shaw KK. Measuring community- university partnerships across a complex research university: lessons and findings from a pilot enterprise data collection mechanism. *Metropolitan Universities* 2015;26(2):124.
20. Zimmerman EB, Raskin SE, Ferrell B, Krist A. Developing a classification system and algorithm to track community-engaged research using IRB protocols at a large research university. *J Clin Transl Sci* 2022;6(1):e6 [FREE Full text] [doi: [10.1017/cts.2021.877](https://doi.org/10.1017/cts.2021.877)] [Medline: [35154815](https://pubmed.ncbi.nlm.nih.gov/35154815/)]
21. Kumar V, Choudhary A, Cho E. Data augmentation using pre-trained transformer models. arXiv Preprint posted online on March 4, 2020. [FREE Full text]
22. Ferrell B. Data-augmentation-for-text-data-nlpaug. GitHub. 2020. URL: <https://github.com/brianferrell787/Data-augmentation-for-text-data-nlpaug> [accessed 2022-08-25]
23. Ma E. nlpaug. GitHub. 2019. URL: <https://github.com/makcedward/nlpaug> [accessed 2022-08-25]
24. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv Preprint posted online on October 2, 2019. [FREE Full text]
25. Schuster M, Paliwal K. Bidirectional recurrent neural networks. *IEEE Trans Signal Processing* 1997;45(11):2673-2681. [doi: [10.1109/78.650093](https://doi.org/10.1109/78.650093)]
26. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997 Nov 15;9(8):1735-1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)] [Medline: [9377276](https://pubmed.ncbi.nlm.nih.gov/9377276/)]
27. Gers F. Long short-term memory in recurrent neural networks. CiteSeerX. 2001. URL: <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.6677> [accessed 2022-08-28]
28. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv Preprint posted online on September 1, 2014. [FREE Full text]
29. Winata GI, Kampman OP, Fung P. Attention-based LSTM for psychological stress detection from spoken language using distant supervision. 2018 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing; April 15-20, 2018; Calgary, AB.
30. Pennington J, Socher R, Manning CD. Glove: global vectors for word representation. 2014 Presented at: Conference on Empirical Methods in Natural Language Processing; October 25-29, 2014; Doha, Qatar.
31. Ferrell B. Interactive-Word-Embedding-Projector-from-Google. GitHub. 2020. URL: <https://github.com/brianferrell787/Interactive-Word-Embedding-Projector-from-Google> [accessed 2022-08-25]
32. Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput* 2019 Jul;31(7):1235-1270. [doi: [10.1162/neco\\_a\\_01199](https://doi.org/10.1162/neco_a_01199)] [Medline: [31113301](https://pubmed.ncbi.nlm.nih.gov/31113301/)]
33. Ferrell B. Visualizing-attention-layer-in-text-classification-deep-learning-algorithm. GitHub. 2020. URL: <https://github.com/brianferrell787/Visualizing-attention-layer-in-text-classification-deep-learning-algorithm> [accessed 2022-08-25]
34. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intelligence Res* 2002 Jun 01;16:321-357. [doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)]
35. Calculate F1 macro in Keras. Knowledge Transfer. 2021 Jun 11. URL: <https://androidkt.com/calculate-f1-macro-in-keras/> [accessed 2022-08-25]
36. Alsentzer E, Murphy JR, Boag W, Weng WH, Jin D, Naumann T, et al. Publicly available clinical BERT embeddings. arXiv Preprint posted online on April 6, 2019. [FREE Full text]
37. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, et al. Unsupervised cross-lingual representation learning at scale. arXiv Preprint posted online on November 5, 2019. [FREE Full text]

38. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a Robustly Optimized BERT Pretraining Approach. arXiv Preprint posted online on July 26, 2019. [FREE Full text]
39. Rajapakse T. Simple Transformers. URL: <https://simpletransformers.ai/> [accessed 2022-08-25]
40. Ferrell B. transformer\_2class Workspace. Weights & Biases. 2020. URL: [https://wandb.ai/brianferrell78/transformer\\_2class?workspace=user-](https://wandb.ai/brianferrell78/transformer_2class?workspace=user-) [accessed 2021-05-15]
41. Biwald L. Experiment tracking with weights and biases. Weights & Biases. 2020. URL: <https://wandb.ai/site> [accessed 2022-05-15]
42. Ferrell B. Classifying-community-engaged-research-with-transformer-based-models. GitHub. 2020. URL: <https://github.com/brianferrell78/Classifying-community-engaged-research-with-transformer-based-models> [accessed 2022-08-25]

## Abbreviations

**BERT:** Bidirectional Encoder Representations From Transformers

**Bi-LSTM:** bidirectional long short-term memory unit

**CEnR:** community-engaged research

**CNN:** convolutional neural network

**DNN:** deep neural network

**GloVE:** Global Vectors for Word Representation

**IRB:** institutional review board

**NLP:** natural language processing

**SMOTE:** Synthetic Minority Oversampling Technique

**VCU:** Virginia Commonwealth University

**XLM-RoBERTa:** Cross-lingual Language Model–Robustly Optimized BERT Pre-training Approach

*Edited by A Mavragani; submitted 30.07.21; peer-reviewed by C Sun, H Li, X Dong; comments to author 10.09.21; revised version received 30.12.21; accepted 15.06.22; published 06.09.22.*

*Please cite as:*

*Ferrell BJ, Raskin SE, Zimmerman EB, Timberline DH, McInnes BT, Krist AH*

*Attention-Based Models for Classifying Small Data Sets Using Community-Engaged Research Protocols: Classification System Development and Validation Pilot Study*

*JMIR Form Res 2022;6(9):e32460*

*URL: <https://formative.jmir.org/2022/9/e32460>*

*doi: [10.2196/32460](https://doi.org/10.2196/32460)*

*PMID: [36066925](https://pubmed.ncbi.nlm.nih.gov/36066925/)*

©Brian J Ferrell, Sarah E Raskin, Emily B Zimmerman, David H Timberline, Bridget T McInnes, Alex H Krist. Originally published in JMIR Formative Research (<https://formative.jmir.org/>), 06.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# “There’s No Heroin Around Anymore. It’s All Fentanyl.” Adaptation of an Opioid Overdose Prevention Counseling Approach to Address Fentanyl Overdose: Formative Study

Vanessa M McMahan<sup>1</sup>, MS, PhD; Justine Arenander<sup>1</sup>, MSW; Tim Matheson<sup>1</sup>, MS, PhD; Audrey M Lambert<sup>2,3</sup>, MPH; Sarah Brennan<sup>1</sup>, BA; Traci C Green<sup>4,5</sup>, MSc, PhD; Alexander Y Walley<sup>2,3</sup>, MSc, MD; Phillip O Coffin<sup>1,6</sup>, MIA, MD

<sup>1</sup>San Francisco Department of Public Health, San Francisco, CA, United States

<sup>2</sup>Grayken Center for Addiction, Clinical Addiction Research and Education Unit, School of Medicine, Boston University, Boston, MA, United States

<sup>3</sup>Boston Medical Center, Boston University, Boston, MA, United States

<sup>4</sup>Heller School for Social Policy and Management, Brandeis University, Waltham, MA, United States

<sup>5</sup>Center of Biomedical Research Excellence on Opioids and Overdose, Rhode Island Hospital, Providence, RI, United States

<sup>6</sup>Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

**Corresponding Author:**

Vanessa M McMahan, MS, PhD

San Francisco Department of Public Health

25 Van Ness Avenue

Suite 500

San Francisco, CA, 94102

United States

Phone: 1 628 217 7469

Email: [vanessa.mcmahan@sfdph.org](mailto:vanessa.mcmahan@sfdph.org)

## Abstract

**Background:** Drug overdose mortality continues to increase, now driven by fentanyl. Prevention tools such as naloxone and medications to treat opioid use disorder are not sufficient to control overdose rates; additional strategies are urgently needed.

**Objective:** We sought to adapt a behavioral intervention to prevent opioid overdose (repeated-dose behavioral intervention to reduce opioid overdose [REBOOT]) that had been successfully piloted in San Francisco, California, United States, to the setting of Boston, Massachusetts, United States, and the era of fentanyl for a full efficacy trial.

**Methods:** We used the assessment, decision, adaptation, production, topical experts, integration, training, and testing (ADAPT-ITT) framework for intervention adaptation. We first identified opioid overdose survivors who were actively using opioids as the population of interest and REBOOT as the intervention to be adapted. We then performed theater testing and elicited feedback with 2 focus groups (n=10) in Boston in 2018. All participants had used opioids that were not prescribed to them in the past year and experienced an opioid overdose during their lifetime. We incorporated focus group findings into our initial draft of the adapted REBOOT intervention. The adapted intervention was reviewed by 3 topical experts, and their feedback was integrated into a subsequent draft. We trained study staff on the intervention and made final refinements based on internal piloting. This paper describes the overall ADAPT-ITT process for intervention adaptation, as well as a qualitative analysis of the focus groups. Working independently, 2 authors (VMM and JA) reviewed the focus group transcripts and coded them for salient and common themes using the constant comparison method, meeting to discuss any discrepancies until consensus was reached. Codes and themes were then mapped onto the REBOOT counseling steps.

**Results:** Focus group findings contributed to substantial changes in the counseling intervention to better address fentanyl overdose risk. Participants described the widespread prevalence of fentanyl and said that, although they tried to avoid it, avoidance was becoming impossible. Using alone and lower opioid tolerance were identified as contributors to overdose risk. Slow shots or tester shots were acceptable and considered effective to reduce risk. Naloxone was considered an effective reversal strategy. Although calling emergency services was not ruled out, participants described techniques to prevent the arrival of police on the scene. Expert review and internal piloting improved the intervention manual through increased participant centeredness, clarity, and usability.

**Conclusions:** We successfully completed the ADAPT-ITT approach for an overdose prevention intervention, using theater testing with people who use opioids to incorporate the perspectives of people who use drugs into a substance use intervention. In the current crisis, overdose prevention strategies must be adapted to the context of fentanyl, and innovative strategies must be deployed, including behavioral interventions.

**Trial Registration:** ClinicalTrials.gov NCT03838510; <https://clinicaltrials.gov/ct2/show/NCT03838510>

(*JMIR Form Res* 2022;6(9):e37483) doi:[10.2196/37483](https://doi.org/10.2196/37483)

## KEYWORDS

opioid overdose; fentanyl; motivational interviewing; naloxone; assessment, decision, adaptation, production, topical experts, integration, training, and testing; ADAPT-ITT; theater testing

## Introduction

### Background

Since 1999, drug overdose deaths have risen 4-fold in the United States, with the majority of overdose deaths related to opioids [1]. We are now in the “third wave” of the overdose epidemic, characterized predominantly by fentanyl and fentanyl-related overdose mortality [2,3]. From 2015 to 2021, the annual number of overdose deaths due to synthetic opioids, which are almost exclusively fentanyl and related analogs, rose 93% from 51,575 to 99,543 [4]. Fentanyl deaths have supplanted heroin deaths; although all opioid deaths have been increasing, heroin-involved deaths that did not also involve fentanyl have decreased [5]. Opioid overdose prevention strategies must be adapted to appropriately address fentanyl overdose.

Providing naloxone to laypeople who may witness an opioid overdose is highly effective at preventing overdose mortality [6-10]; however, the rapidity of fentanyl overdose limits the time frame in which naloxone can be effectively administered. Therefore, additional and complementary opioid overdose interventions are urgently needed [11]. Motivational interviewing has been shown to reduce opioid risk behaviors [12,13]. From 2014 to 2016, we completed a pilot in San Francisco, California, United States, of a repeated-dose motivational interviewing intervention (repeated-dose behavioral intervention to reduce opioid overdose [REBOOT]) aimed at reducing overdose occurrence and mortality among persons who had previously overdosed and had received take-home naloxone. Participants randomized to the REBOOT intervention were less likely to experience an opioid overdose during 16 months of follow-up, and those who did overdose had fewer overdose events [14]. Building on these pilot results, we designed a clinical trial (REBOOT 2.0) to determine the efficacy of the REBOOT intervention on overdose occurrence and number of overdoses (NCT03838510).

### Adapting the REBOOT Counseling Intervention

To address the emergence of fentanyl in eastern US states between the pilot and REBOOT 2.0, as well as to ensure that the intervention was appropriate for a broader geographic region, we sought to adapt the REBOOT counseling intervention using the assessment, decision, adaptation, production, topical experts, integration, training, and testing (ADAPT-ITT) model [15]. We describe the ADAPT-ITT process and focus group findings, highlighting how the perspectives of people who use drugs

guided the intervention adaptation process to address fentanyl overdose risk.

## Methods

### ADAPT-ITT Framework

ADAPT-ITT is an effective framework for adapting HIV-prevention evidence-based interventions to new populations and settings, which includes 8 sequential phases to iteratively elicit feedback from the population of interest and key stakeholders [15]. This is the first study to our knowledge to use ADAPT-ITT to adapt an opioid overdose prevention intervention, although it has been used previously to adapt other interventions for people who use opioids to new settings [16,17]. We conducted a modified ADAPT-ITT process to adapt the REBOOT overdose counseling intervention to be appropriate in the context of fentanyl.

### Phases 1 (Assessment) and 2 (Decision)

The aim of phase 1 of the ADAPT-ITT model is to identify the population at risk, and the aim of phase 2 is to identify the appropriate intervention to adapt. We identified people with overdose history as the population at risk, considering their elevated risk for opioid overdose [18], and chose to adapt the REBOOT counseling intervention based on its success during our pilot [14]. Therefore, we modified phases 1 and 2 of the ADAPT-ITT approach to be internal, and we did not review additional overdose reduction interventions for potential adaptation. As these phases were completed before project initiation, they are not described further herein.

### REBOOT Counseling Intervention

The REBOOT counseling intervention aimed to engage participants in a discussion of their opioid overdose history, both witnessed and experienced; discuss overdose risk behaviors; and identify feasible risk reduction strategies that the participant would be interested in using. Specifically, the 6 counseling steps are as follows: (1) explore participants' experiences witnessing an overdose and contributing factors, (2) discuss participants' own most recent overdose and contributing factors, (3) review overdose risk behaviors, (4) review ways to reduce overdose risk, (5) discuss participants' current and additional planned strategies to reduce overdose risk and provide them with a wallet card listing risk-reduction strategies, and (6) review 3 steps in the Skills and Knowledge on Overdose Prevention (SKOOP) Project overdose recognition and reversal curriculum (ie, recognize an overdose, respond to

an overdose, and provide aftercare) [19]. In addition, during the session, the counselor provides and reviews with participants 3 handouts that describe activities that increase opioid overdose risk, ways to reduce risk, and steps to respond to a witnessed overdose. REBOOT is intended for individuals who have previously received take-home naloxone, but we do not distribute naloxone as part of the counseling intervention.

### Phase 3 (Administration)

#### Overview

In phase 3 of ADAPT-ITT, the intervention is theater tested, and feedback from the audience is elicited. Theater testing involves inviting individuals of the population at risk to watch a demonstration of the intervention and then answer questions about their opinions of the intervention. On September 26, 2018, we conducted theater testing with 2 focus groups in Boston, Massachusetts, United States. REBOOT had been piloted in San Francisco, and at that time there was very little fentanyl being used in that city; thus, the adaptation effort was focused on Boston, where fentanyl was the dominant opioid [20-22].

#### Focus Group Participants

We contacted individuals who had reported opioid use while participating in previous studies at the Boston Medical Center and provided consent for future contact for research. Eligibility criteria for the focus groups included age 18 to 65 years, living in the Boston metropolitan area, using an opioid in the past year that was not prescribed, lifetime history of overdose, and comfort participating in the focus group in English. Participants who were screened for eligibility reported their demographic characteristics (ie, age, gender, race, and ethnicity), overdose history, lifetime naloxone use, and past-year opioid use. Participants reviewed an information sheet with study staff and provided verbal consent for participation.

#### Focus Group Procedures

In each focus group, participants were shown the same video of a counseling session with a participant from San Francisco. The counseling session lasted 39 minutes and was conducted by a trained clinical psychologist with expertise in motivational interviewing following the REBOOT counseling steps. The handouts provided to the participant at the counseling session were also provided to focus group participants for review. A video recording of the session was used, rather than a live demonstration, because of the short-acting nature of fentanyl and the risk that focus group participants would develop opioid withdrawal symptoms if there were delays in the procedures.

During the viewing of the video and afterward, participants were asked questions following a semistructured interview guide about their opinions and perspectives regarding the counseling session (Multimedia Appendix 1). At the end of each focus group, participants were given US \$40 for their participation. Focus groups were audio recorded and transcribed using Rev software [23].

### Focus Group Analysis

Working independently, 2 authors (VMM and JA) reviewed the focus group transcripts and coded them for salient and common themes using the constant comparison method [24]. A list of a priori codes of common themes related to opioid overdose was used to help guide transcript coding (eg, “access to naloxone,” “tester shot,” and “calling 911”). VMM and JA met after coding each focus group transcript to compare codes and discuss them until they reached consensus. As codes emerged as focus group transcripts were reviewed, both focus group transcripts were reviewed a second time after the final list of codes was determined. Codes and themes were then mapped onto the 6 steps of the REBOOT counseling intervention. All transcript coding and analyses were performed using ATLAS.ti software (version 8; ATLAS.ti Scientific Software Development GmbH).

### Ethics Approval

This study was approved by the University of California San Francisco Institutional Review Board (17-24203).

### Phases 4 (Production), 5 (Topical Experts), and 6 (Integration)

Focus group findings were used to produce a draft of the adapted counseling intervention (phase 4), which was reviewed by 3 topical experts (phase 5). The experts were overdose education and prevention professionals from the Boston Public Health Commission, Harm Reduction Coalition, and the Education Development Center. The expert feedback was integrated into the subsequent draft of the counseling intervention (phase 6).

### Phases 7 (Training) and 8 (Testing)

We then trained our counseling staff (phase 7) and tested the adapted intervention (phase 8) to make final refinements. We modified phase 8 of ADAPT-ITT by conducting internal monitoring with the study team and not a pilot with the population of interest. The final adapted counseling intervention is being evaluated in the ongoing REBOOT clinical trial.

## Results

### Phase 3: Focus Group Findings

In total, 21 individuals were eligible and invited to participate in the focus groups. Of these 21 individuals, only 10 (48%) attended the focus groups, with 6 (60%) in one focus group and 4 (40%) in the other. The mean age of participants was 44.1 (SD 13.8) years, and the majority were men (7/10, 70%). Half (5/10, 50%) of the sample was Hispanic/Latinx; in terms of race, 20% (2/10) were Black or African American, 30% (3/10) were White or Caucasian, and 50% (5/10) reported another race. All participants had experienced an overdose involving fentanyl, and 90% (9/10) had used fentanyl in the past year (Table 1).

**Table 1.** Participant demographic characteristics and experience with fentanyl (N=10).

Demographic characteristics	Values
<b>Number of participants, n (%)</b>	
Focus group 1	6 (60)
Focus group 2	4 (40)
Age (years), mean (SD)	44.1 (13.8)
<b>Ethnicity, n (%)</b>	
Non-Hispanic/Latinx	5 (50)
Hispanic/Latinx	5 (50)
<b>Race, n (%)</b>	
Black or African American	2 (20)
White or Caucasian	3 (30)
Other race	5 (50)
<b>Gender, n (%)</b>	
Male	7 (70)
Female	3 (30)
<b>Fentanyl, n (%)</b>	
Used fentanyl in the past year	9 (90)
Experienced an overdose involving fentanyl	10 (100)

### Step 1: Explore Participants' Experiences Witnessing an Opioid Overdose, Factors Contributing to This Overdose, and Thoughts About Own Substance Use Since This Witnessed Overdose

All participants reported that they had witnessed an opioid overdose. Participants described various strategies to respond to a witnessed overdose, including using naloxone, calling 911, and other strategies stemming from their lived experience.

#### Naloxone

All participants had previously received take-home naloxone and were generally knowledgeable about how to use naloxone. In total, 80% (8/10) of the participants reported administering naloxone to someone who was overdosing:

*Narcan [a brand name for naloxone]...I have that in my house for my son. I have a ring and it shows you what to do... [Focus group 1, attendee 2]*

#### Calling 911

Participants thought that calling 911 had inherent risks, including arrest and loss of housing. Strategies to mitigate these risks included relying on the Good Samaritan law, although there was discussion regarding how it only protected the individual calling 911 from arrest, and not others present at the scene of an overdose. Participants also recommended not telling a dispatcher that the person had overdosed, but instead saying that they were "not breathing" or "unresponsive" so that medical services would arrive without the police:

*A lot of people are afraid to call the cops [for an overdose] because they think they're going to get*

*locked up. I've seen people like, "Leave him, leave him. Let's go." [Focus group 1, attendee 5]*

*I'm not trying to be a jerk, but you pull [the person who overdosed] out in the hallway and [tell 911], "Listen. I came outside, and there's somebody unconscious on my ground. Is it drug-related? I don't know. Here's the address." And hang up quick so they can make it there quick. [Focus group 1, attendee 1]*

#### Additional Overdose Response Strategies

Participants described several strategies to reverse a witnessed opioid overdose that stemmed from their lived experience. These included injecting the individual who had overdosed with salt and water or cocaine and placing the individual in an ice bath or putting ice on their genitals, which participants emphasized they had seen be successful:

*They shoot you with salt. They put ice on your testicles. Water...They just inject the salt, and it gives you a reaction without being in the vein, then it wakes you up. I have seen it with my own eyes, people got woken up with a shot of salt. They say it doesn't work but I have seen it work. [Focus group 1, attendee 4]*

### Step 2: Discuss Participants' Own Most Recent Overdose and Contributing Factors

All focus group participants reported experiencing an opioid overdose during their lifetime, consistent with eligibility criteria, and all participants reported a lifetime overdose that involved fentanyl. Among those who described their overdose experiences, most had been administered naloxone to reverse the overdose. A lower opioid tolerance after being released from



jail or leaving treatment was commonly identified as a contributing factor for opioid overdose:

*I had just come out of jail three weeks prior. I was already using dope but what I was doing was I was using tester shots. If I had a 30 for myself, I would do a 15 or like a 10 [to] see how it made me feel. If it made me feel alright, I would just save the other shot, because, at the end of the day, you can only just do more. I looked at my girlfriend, I'm like "Fuck that I'm going to do the whole 30." I did it and I felt the heat from my toes up I was like "Damn that shit was strong." Next thing I remember, I'm holding onto a fence, I'm like "What the hell?" I'm looking around, she's crying. I'm like "What the hell happened?" She's like you've been dead for like seven minutes. I'm like "Oh my god." She had just woke me up. They Narcan-ed me like three times, sternum rubbed me and when they threw water on me, I guess that's when I responded.* [Focus group 1, attendee 6]

### Step 3: Review Behaviors That Increase Opioid Overdose Risk

#### Overview

In step 3 of the counseling session, the counselor provided a handout to the participant listing several behaviors that increase the risk for an overdose ([Multimedia Appendix 2](#)). The counselor read the list out loud with the participant and then explored the participant's own risk for overdose. After viewing the counseling session and reading the handout, the focus group participants highlighted 3 main risk factors for opioid overdose: fentanyl, opioid tolerance, and duration of opioid use.

#### Fentanyl

Participants across both focus groups described fentanyl as prevalent and a major contributor to opioid overdose:

*They [are] even putting [fentanyl] on the weed. On the cocaine...* [Focus group 1, attendee 4]

*There's no heroin around anymore. It's all fentanyl.* [Focus group 2, attendee 4]

Participants also described how fentanyl was being mixed with other drugs unbeknownst to those using those drugs:

*They're putting [fentanyl] in the coke now, the cocaine. My friend [who] only does coke...when he stopped doing coke, he was like dope sick. He don't do dope or nothing and had all the symptoms of dope sickness and then—it's a couple—when she did her test at the methadone clinic, she was coming up for fentanyl and she was only doing coke.* [Focus group 1, attendee 2]

#### Opioid Tolerance

Lower opioid tolerance after not using opioids for a period of time, in particular after being in jail or substance use disorder treatment, was identified as a risk factor for overdose:

*Every time someone just get[s] out [of jail or treatment] they think they can do the same thing as*

*when they went in...[but] they'll automatically OD.* [Focus group 2, attendee 1]

#### Duration of Opioid Use

Focus group participants described overdose risk as higher among people who had recently begun using opioids. Participants frequently described themselves as people who had used opioids for a long period of time and therefore were at less risk of overdose. The median length of time since first using opioids that were not prescribed was 26.5 (IQR 17-35) years. A participant described the utility of counseling among those starting to use opioids:

*I think [overdose counseling] would be helpful for kids who are just starting using, who aren't like smart into the lifestyle and know the game...who are more reckless than a lot of us who have been using for years.* [Focus group 2, attendee 4]

Although participants described their duration of opioid use as lowering their overdose risk during the focus groups, on the screening questionnaire, of the 10 participants, 7 (70%) said that they felt their risk of overdose was high, 2 (20%) said that it was moderate, and 1 (10%) reported low risk. The participant who reported low overdose risk had used opioids for the shortest period of time (5 years). Thus, although participants who had used opioids longer thought that their overdose risk was lower than people starting to use opioids, they still reported a moderate to high risk for overdose.

### Step 4: Review Ways to Reduce Overdose Risk

In step 4 of the counseling session, the counselor provided another handout to the participant that listed several strategies that could be used to reduce one's risk of opioid overdose ([Multimedia Appendix 2](#)). After watching the counseling session and reading the corresponding handout, focus group participants discussed 5 main strategies to reduce the risk of opioid overdose: not using alone, avoiding fentanyl, using the same dealer, using less or using partial shots or tester shots, and medications for opioid use disorder (MOUD).

#### Do Not Use Alone

There was consensus that not using alone was an effective way to lower the risk of opioid overdose, and most participants reported that they usually use with others, and when they use alone, they may call someone, use in a bathroom in a monitored environment (eg, community-based organization), or make sure a friend can access them:

*I try to hang out with friends, I don't use by myself now. It's like you kind of do it in a group, you know? So, each know each other's tolerances.* [Focus group 1, attendee 3]

*When I do [opioids] by myself in my house, I leave the doors unlocked, you know? I got a good friend of mine that has a copy of a key. That's the only thing that I do [for overdose prevention].* [Focus group 2, attendee 5]

### **Avoid Fentanyl**

Across both focus groups, participants discussed how they aimed to avoid fentanyl because of its prevalence and potency; however, they considered avoiding fentanyl increasingly difficult because of the rapidly rising prevalence of fentanyl. Ways to identify fentanyl included its gray color; sweet taste; and smell, which is different from that of heroin. Using fentanyl test strips, as suggested in the counseling video, were discussed but were not considered accessible to focus group participants in Boston at the time:

*[I do] a whole lot [differently because of fentanyl in the drug supply], make sure you don't get that bullshit. Go to somebody you know that's got real dope.* [Focus group 2, attendee 3]

*Yeah, a lot of them [lie and] say "No, it's straight dope, it's straight dope. There's no fentanyl." That's why a lot of people they're doing pills. They rather be doing pills than do the dope.* [Focus group 1, attendee 6]

Participants also described a paradox, whereby although they typically tried to avoid fentanyl, if they learned of strong opioids, they wanted to use them—they sought them out:

*Some people would see somebody overdose and be like, yo, who did he buy that from?...Because if it puts you out then it means it's some good shit. So other people be like, yo, where do you buy that from? I want some of that shit.* [Focus group 1, attendee 3]

*If over there in the corner, somebody comes and tells me..."that shit over there is killing motherfuckers," I'm going to go over there, excuse my language, because there's some good dope over there. We never learn. You understand? I'll ask who's got the good stuff, "that one's killing people be careful" and I'll go over there.* [Focus group 1, attendee 4]

One way that participants rationalized this paradox was that they thought that they personally would not overdose from drugs that others had overdosed from because they had a higher tolerance:

*We don't think we're going to OD off of it. We think like, "Oh, my habit's worse than your habit..." I have a 500-dollar-a-day habit.* [Focus group 2, attendee 4]

### **Using Less or Using Partial Shots or Tester Shots**

Participants discussed strategies to reduce the amount of drugs used or sample drugs before using them, through either partially depressing the plunger before injecting all of the solution (*slow shot*) or doing a tester shot. Participants stressed that using less was especially important when fentanyl might be present:

*If you know there's fentanyl in there, instead of doing a 40 you do a 20...It will be harm reduction for myself because if I'm doing a 40 of fentanyl now, and then I come and test it before I do that 40 and it's got fentanyl, I'm going to do a 20 because the 40 is going to kill me.* [Focus group 1, attendee 4]

However, there were mixed feelings about partial shots or tester shots, and positive perceptions were generally shared in the context of expected lower tolerance (eg, having recently been incarcerated):

*When I'm getting [high], I'm getting [high]. Forget about that partial test.* [Focus group 1, attendee 4]

*I've learned that every time I get out of jail I just do less, like a tester shot now. My boyfriend would always say to me, you can always do more, but you can't take one back. So, stop being a glutton, it's right here...It stuck in my head like he's got a point. You can do another one, but you can't take one back.* [Focus group 2, attendee 2]

### **Use the Same Dealer**

The participant in the counseling video used the same dealer to avoid getting fentanyl. Across both focus groups, participants did not think that using the same dealer was an effective way to avoid fentanyl because of the uncertainty in the drug supply:

*[The lady in the video] gets it from the same guy all the time...That don't mean nothing. The guy doesn't know what chemicals the drug got. He could say, "It's good don't worry about it," and when she gets home [she overdoses]. You never know what could happen.* [Focus group 1, attendee 4]

*Well, because he could be the same drug dealer, but he might have gotten a different batch. They don't always get the same thing. There are different grades of fentanyl. There are stronger grades that have legs, there are shit grades that you really get pissed about...* [Focus group 2, attendee 2]

### **Treatment for Opioid Use Disorder**

Although, in line with the participant-centered harm reduction counseling approach, MOUD was included on the handout of possible strategies to reduce the risk of opioid use disorder, it would only be discussed with the participant if they demonstrated interest in MOUD. A few (2/10, 20%) of the focus group participants said that they thought this was not sufficient and that the counseling intervention would be more effective if MOUD was emphasized more:

*I think it was helpful to encourage [the participant in the video to stay] clean, but [the counselor in the video] didn't offer any places like we should try, like methadone maintenance, if you haven't been able to maintain completely off anything, or Suboxone or halfway house.* [Focus group 2, attendee 2]

### **Step 5: Create or Provide a Wallet Card for Participants Based on Their Current and New Efforts to Reduce Risk of Overdose**

Toward the end of the counseling session, the counselor created a wallet card for the participant based on the overdose prevention strategies that the participant described as part of their current practice or additional strategies of interest to them. Focus group participants thought that providing a wallet card to participants to remind them of what was discussed in the

counseling session was helpful. Using other forms of technology to remind participants of the discussed strategies was also suggested:

*Maybe even, nowadays with technology, it could even be on an iPad...Or text messages, or another app that you can check on.* [Focus group 1, attendee 2]

Although participants thought that the wallet card was useful, participants recommended providing more handouts, naloxone, and fentanyl test strips, as well as providing training for rescue breathing and naloxone administration:

*If you're actively using, [a brochure about resources is] just a piece of paper, it's nothing, you know what I mean? But if you're trying to change your life and get help then like it can be extremely helpful. But it all depends on where you are.* [Focus group 2, attendee 4]

### Step 6: Review 3 Steps in Management of Witnessed Overdose (SKOOP Curriculum)

The last step was to review a handout that described the 3 steps in management of a witnessed overdose from the SKOOP Curriculum [19]: recognizing an overdose, responding to an overdose, and providing aftercare. Focus group participants commented on how the participant in the video learned additional information about overdose response during this step and that the use of a handout was helpful:

*When [the counselor in the video] had it in black and white right in front of her and she was reading and she's like, "Oh I didn't know about the recovery position," so she was very perceptive to it, you know what I mean?...rather than just talking about things, like when it's in black and white, right in front of you. Like for me, I tend to pick things up by reading them...'cause I'll forget what you're telling me...if it's right in front of me, it's hard to ignore.* [Focus group 2, attendee 4]

### Overall Suggestions for the Counseling Session

Focus group participants reported that they already knew a lot of the information that was shared in the counseling session, and they thought that the best audience would be younger people who have not used opioids for very long:

*[The counseling session] might be helpful for the people just getting into this game, but I've been in this game since '69, so it might be helpful to the newcomers.* [Focus group 2, attendee 1]

Some (3/10, 30%) of the participants recommended including a conversation about the root causes of people's substance use and expanding the intervention to include a broader counseling intervention:

*I've heard [the content of the counseling session] multiple times. I think I agree with the whole overdose prevention and training in Narcan, reiterating it and whatnot. But I think they could've focused on a little bit more of core issues of why people use and why do they continue to go back. Because [the woman in the*

*video] had six months clean in jail. You know what I mean? Like she didn't have to go back and get high. Like granted nine out of 10 of us are going to. You know what I mean? But like at that point you're already over the physical part of it. So, like start dealing with your mental side and like those cravings...I think they could like integrate that into a lot of the conversations.* [Focus group 2, attendee 4]

### Phase 4: Production

We incorporated the focus group findings to adapt the REBOOT counseling intervention to better address fentanyl overdose risk. Specifically, we highlighted avoidance of fentanyl and using tester shots and de-emphasized using the same dealer. We added the recommendation to test drugs for fentanyl and identified locations to refer participants to obtain fentanyl test strips. Counseling staff were alerted that even among those who are aware of fentanyl's overdose risk there may be a desire to use fentanyl; they were directed to explore this paradox with individuals who use opioids during counseling and tailor overdose prevention strategies as appropriate. Staff were also trained to understand that participants may underestimate their overdose risk, and they were provided guidance on how to help participants develop a realistic personal overdose risk assessment. We also added tips when calling 911 to mitigate the chances of police arriving at the overdose scene.

### Phases 5 and 6: Topical Experts and Integration

We shared a draft of the counseling manual describing the adapted REBOOT intervention informed by the focus groups with 3 overdose prevention and education experts for their review and feedback. All 3 experts sent suggestions for the draft intervention, which were incorporated into the next iteration of the counseling approach.

Recommendations included adding less risky routes of administration as a way to reduce overdose risk (eg, smoking instead of injecting) and asking participants to "walk through" a typical day to visualize how overdose prevention strategies could realistically be implemented in their life, as well as asking participants what they want to "change" about their substance use, not whether they want to "reduce" or "stop" their substance use.

In response to participant and expert feedback, we developed an assessment of the level of interest in substance use treatment and any plans or needs to access substance use treatment, which is conducted in at least one of the 4 REBOOT counseling sessions. Similar to the overdose counseling approach, the substance use treatment counseling component aims to be participant-centered and begins by exploring participants' previous substance use treatment experience before assessing interest.

### Phases 7 and 8: Training and Testing

After incorporating the expert feedback into the counseling intervention, we trained study site staff by reviewing the manual and handouts, and team members conducted role plays with a trained clinical psychologist. Study staff provided further minor modifications to the approach to best operationalize it.

Recommendations included improving readability (eg, using bullets, larger font, and more spacing), including the expected duration of the counseling session at the beginning, and using a more open-ended question when assessing the participants' personal overdose history (ie, "Now let's discuss the last time you experienced an overdose. Thinking back to that time, please tell me the story of your most recent overdose").

## Discussion

### Findings and Contextualization

We adapted an opioid overdose prevention intervention to the era of fentanyl and a broader geographic region through the ADAPT-ITT process. As expected, the prevalence of fentanyl in Boston was the dominant issue for overdose prevention, and the process resulted in multiple adaptations to the final intervention. Focus group participants who used opioids emphasized the importance of fentanyl avoidance (eg, by assessing drug color, taste, and smell), as well as the growing impossibility of avoiding fentanyl. The predominant role of fentanyl in increasing overdose risk has also been reported in other studies among people who use opioids [25-27].

Although participants described the importance and difficulty of avoiding fentanyl, they also described the desire to use drugs that were involved in an overdose because of that indicated potency. This mix of fentanyl avoidance and seeking was also seen in a qualitative study in Pennsylvania, United States, among persons with recent opioid misuse or heroin use [28]. Participants seemed to be aware of this paradox but rationalized it by characterizing themselves as using opioids for a long time and having a high opioid tolerance and, therefore, at lower risk of an overdose. This was similar to our findings among REBOOT pilot participants, where participants who were older had lower odds of perceived overdose risk [29]. Pilot participants similarly described inexperience as a main contributor to overdose events that they witnessed in other people but not to overdoses that they personally experienced [30]. This finding led to a greater emphasis in the intervention on helping participants develop a realistic assessment of their risk for overdose.

The predominant role of fentanyl in opioid overdose affected participants' perceptions of traditional overdose prevention strategies. Using the same dealer was not considered an effective strategy to reduce overdose risk because of the uncertainty in the drug supply at every level, which has been reported in another study among people who have survived opioid overdose [26]. Tester shots or slow shots were acceptable and considered particularly important if fentanyl could be a possible contaminant and after periods of not using opioids (eg, incarceration). Fentanyl test strips were generally viewed as helpful but were considered unavailable at the time we conducted the focus groups. Studies have shown fentanyl test strips to be acceptable, easy to use, and associated with overdose prevention behaviors (eg, not using drugs that test positive for fentanyl and having naloxone available) [31]. As the lack of a relative safety benefit from smoking fentanyl, compared with injecting it, has become apparent, we have tailored risk reduction messages regarding route of administration to the opioid being

used, and we do not explicitly recommend smoking fentanyl for overdose prevention.

In addition to opioid overdose prevention strategies, we discussed effective overdose response. Participants had experience responding to overdose, including using naloxone, which was considered an effective and acceptable response strategy. Participants also described witnessing overdose reversals when stimulation was used (eg, applying something cold). Study participants considered these alternative response strategies as effective, which has been reported in another study among people who use opioids [32]. For naloxone or other overdose response strategies to be effective, there must be someone present who is able to respond. Of the 10 participants, 3 (30%) reported using opioids alone, and they described strategies to remain accessible to others in the event that they experienced an overdose. In a study among people who recently used nonmedicinal opioids in New York City, New York, United States, age of  $\geq 50$  years and non-Hispanic Black race were associated with increased odds of not having naloxone present and a person trusted to administer it when using opioids [33]. Messages highlighting the importance of not using opioids alone should be emphasized among individuals who are less likely to have someone present if they overdose, which may include older and non-Hispanic Black people who use opioids.

There was a common fear of arrest or other negative consequences related to calling 911 in the event of an overdose, a common finding from prior research [25,34-41] that does not seem to have been significantly mitigated by Good Samaritan legislation. Techniques to minimize negative outcomes were discussed by participants, including telling the dispatcher that the person was "not breathing" and not that they had used drugs, which aligns with overdose response guidance from the National Harm Reduction Coalition [42]. Providing individuals with information about what to communicate to dispatchers may alleviate some fears associated with calling 911.

The subsequent phases of the ADAPT-ITT process, including expert review and testing, further helped to refine the intervention. These later modifications were typically minor, including additional detail to improve clarity and changing the manual's formatting to increase usability. After internal testing at the study sites, we finalized the REBOOT 2.0 counseling manual and are currently evaluating the efficacy of the intervention in our ongoing clinical trial.

### Limitations

This study includes limitations. The focus groups were small, which can limit the generalizability of findings; however, considering the narrow and clear scope of the research (ie, adapting an existing intervention), active participation of the attendees, and repeat themes that arose during conversation, we believe that the number of attendees ( $n=10$ ) was sufficient for the aims of this study [43]. Our findings may have been affected by social desirability and selection biases [40,44]. Nonetheless, our findings are similar to those among other groups of people who use opioids [29,30]. The study was conducted in a single setting, limiting the generalizability of findings to other locations. However, theater testing in the ADAPT-ITT framework is intended to be conducted among the population

that the intervention is being adapted for, which in this case was survivors of opioid overdose who use opioids in Boston [15]. As fentanyl has since become more prevalent across the United States, our findings may have become relevant to other locations across the country.

Of the 10 participants, 3 (30%) were women, and additional research may be needed to further explore potential gender-specific intervention adaptations [39,45]. We requested feedback from 3 topical experts and our study teams, and their opinions may not reflect those of other experts and service providers of people who use opioids. Our focus group interview guide asked broad questions about participants' perceptions of the counseling intervention and did not elicit feedback for each step of the intervention. More targeted questions could have provided more detailed recommendations for how to improve each counseling step.

## Conclusion

We successfully used the ADAPT-ITT approach to modify an opioid overdose prevention intervention, ensuring its applicability in a new geographic area and in the setting of a more potent street opioid. We conducted theater testing with people who use opioids, a challenge given the short-acting nature of fentanyl. Important adaptations resulted from this stage of the process, which led to a more appropriate intervention for people who use fentanyl. Incorporating input and lessons learned from people who use substances was key to optimizing our counseling intervention and messaging. Additional improvements were made through expert review and internal piloting. As fentanyl became the dominant street opioid in San Francisco during the full REBOOT trial, these adaptations proved essential.

---

## Acknowledgments

This study was supported by the National Institute on Drug Abuse (R01 DA037194). The funding agency had no role in study design, data collection, analysis, the decision to publish, or the preparation of the manuscript. The authors are thankful for the time and insights of their focus group participants. The authors also thank Amy Delaney, Maya Doe Simkins, and Eliza Wheeler for their expertise and input.

---

## Authors' Contributions

VMM and JA conducted data analysis. TM participated as the counselor in the theater-tested counseling session. JA, AML, TCG, AYW, and POC observed or facilitated focus groups. JA, TM, AML, SB, TCG, AYW, and POC contributed to intervention development. VMM led manuscript development. All authors reviewed, and provided feedback on, the manuscript.

---

## Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Focus group guide.

[\[DOCX File, 26 KB - formative\\_v6i9e37483\\_app1.docx\]](#)

---

### Multimedia Appendix 2

Focus group handouts.

[\[PDF File \(Adobe PDF File\), 156 KB - formative\\_v6i9e37483\\_app2.pdf\]](#)

---

## References

1. Understanding the Epidemic. Centers for Disease Control and Prevention. 2021 Mar 17. URL: <https://www.cdc.gov/drugoverdose/epidemic/index.html> [accessed 2021-05-04]
2. Ciccarone D. The triple wave epidemic: supply and demand drivers of the US opioid overdose crisis. *Int J Drug Policy* 2019 Sep;71:183-188 [FREE Full text] [doi: [10.1016/j.drugpo.2019.01.010](https://doi.org/10.1016/j.drugpo.2019.01.010)] [Medline: [30718120](https://pubmed.ncbi.nlm.nih.gov/30718120/)]
3. Zoorob M. Fentanyl shock: the changing geography of overdose in the United States. *Int J Drug Policy* 2019 Aug;70:40-46. [doi: [10.1016/j.drugpo.2019.04.010](https://doi.org/10.1016/j.drugpo.2019.04.010)] [Medline: [31079029](https://pubmed.ncbi.nlm.nih.gov/31079029/)]
4. Ahmad FB, Cisewski JA, Rossen LM, Sutton P. Provisional drug overdose death counts. National Center for Health Statistics. 2022. URL: <https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm> [accessed 2022-02-17]
5. Mattson CL, Tanz LJ, Quinn K, Kariisa M, Patel P, Davis NL. Trends and geographic patterns in drug and synthetic opioid overdose deaths - United States, 2013-2019. *MMWR Morb Mortal Wkly Rep* 2021 Feb 12;70(6):202-207 [FREE Full text] [doi: [10.15585/mmwr.mm7006a4](https://doi.org/10.15585/mmwr.mm7006a4)] [Medline: [33571180](https://pubmed.ncbi.nlm.nih.gov/33571180/)]
6. Bennett AS, Bell A, Tomedi L, Hulsey EG, Kral AH. Characteristics of an overdose prevention, response, and naloxone distribution program in Pittsburgh and Allegheny County, Pennsylvania. *J Urban Health* 2011 Dec;88(6):1020-1030 [FREE Full text] [doi: [10.1007/s11524-011-9600-7](https://doi.org/10.1007/s11524-011-9600-7)] [Medline: [21773877](https://pubmed.ncbi.nlm.nih.gov/21773877/)]

7. Walley AY, Xuan Z, Hackman HH, Quinn E, Doe-Simkins M, Sorensen-Alawad A, et al. Opioid overdose rates and implementation of overdose education and nasal naloxone distribution in Massachusetts: interrupted time series analysis. *BMJ* 2013 Jan 30;346:f174 [FREE Full text] [doi: [10.1136/bmj.f174](https://doi.org/10.1136/bmj.f174)] [Medline: [23372174](https://pubmed.ncbi.nlm.nih.gov/23372174/)]
8. Maxwell S, Bigg D, Stanczykiewicz K, Carlberg-Racich S. Prescribing naloxone to actively injecting heroin users: a program to reduce heroin overdose deaths. *J Addict Dis* 2006;25(3):89-96. [doi: [10.1300/J069v25n03\\_11](https://doi.org/10.1300/J069v25n03_11)] [Medline: [16956873](https://pubmed.ncbi.nlm.nih.gov/16956873/)]
9. Lankenau SE, Wagner KD, Silva K, Kecojevic A, Iverson E, McNeely M, et al. Injection drug users trained by overdose prevention programs: responses to witnessed overdoses. *J Community Health* 2013 Feb;38(1):133-141 [FREE Full text] [doi: [10.1007/s10900-012-9591-7](https://doi.org/10.1007/s10900-012-9591-7)] [Medline: [22847602](https://pubmed.ncbi.nlm.nih.gov/22847602/)]
10. Wheeler E, Jones TS, Gilbert MK, Davidson PJ, Centers for Disease Control and Prevention (CDC). Opioid overdose prevention programs providing naloxone to laypersons - United States, 2014. *MMWR Morb Mortal Wkly Rep* 2015 Jun 19;64(23):631-635 [FREE Full text] [Medline: [26086633](https://pubmed.ncbi.nlm.nih.gov/26086633/)]
11. Coffin PO, Maya S, Kahn JG. Modeling of overdose and naloxone distribution in the setting of fentanyl compared to heroin. *Drug Alcohol Depend* 2022 Jul 01;236:109478. [doi: [10.1016/j.drugalcdep.2022.109478](https://doi.org/10.1016/j.drugalcdep.2022.109478)] [Medline: [35588609](https://pubmed.ncbi.nlm.nih.gov/35588609/)]
12. Strang J, McDonald R, Alqurshi A, Royall P, Taylor D, Forbes B. Naloxone without the needle - systematic review of candidate routes for non-injectable naloxone for opioid overdose reversal. *Drug Alcohol Depend* 2016 Jun 01;163:16-23. [doi: [10.1016/j.drugalcdep.2016.02.042](https://doi.org/10.1016/j.drugalcdep.2016.02.042)] [Medline: [26996745](https://pubmed.ncbi.nlm.nih.gov/26996745/)]
13. Bohnert AS, Bonar EE, Cunningham R, Greenwald MK, Thomas L, Chermack S, et al. A pilot randomized clinical trial of an intervention to reduce overdose risk behaviors among emergency department patients at risk for prescription opioid overdose. *Drug Alcohol Depend* 2016 Jun 01;163:40-47. [doi: [10.1016/j.drugalcdep.2016.03.018](https://doi.org/10.1016/j.drugalcdep.2016.03.018)] [Medline: [27062245](https://pubmed.ncbi.nlm.nih.gov/27062245/)]
14. Coffin PO, Santos GM, Matheson T, Behar E, Rowe C, Rubin T, et al. Behavioral intervention to reduce opioid overdose among high-risk persons with opioid use disorder: a pilot randomized controlled trial. *PLoS One* 2017 Oct 19;12(10):e0183354 [FREE Full text] [doi: [10.1371/journal.pone.0183354](https://doi.org/10.1371/journal.pone.0183354)] [Medline: [29049282](https://pubmed.ncbi.nlm.nih.gov/29049282/)]
15. Wingood GM, DiClemente RJ. The ADAPT-ITT model: a novel method of adapting evidence-based HIV interventions. *J Acquir Immune Defic Syndr* 2008 Mar 01;47 Suppl 1:S40-S46. [doi: [10.1097/QAI.0b013e3181605df1](https://doi.org/10.1097/QAI.0b013e3181605df1)] [Medline: [18301133](https://pubmed.ncbi.nlm.nih.gov/18301133/)]
16. Staton M, Knudsen HK, Walsh SL, Oser C, Pike E, Lofwall M. Adaptation of a standard extended-release naltrexone (XR-NTX) protocol for rural re-entering offenders with OUD. *Health Justice* 2021 Feb 05;9(1):4 [FREE Full text] [doi: [10.1186/s40352-021-00130-0](https://doi.org/10.1186/s40352-021-00130-0)] [Medline: [33547578](https://pubmed.ncbi.nlm.nih.gov/33547578/)]
17. Cochran G, Gordon AJ, Field C, Bacci J, Dhital R, Ylioja T, et al. Developing a framework of care for opioid medication misuse in community pharmacy. *Res Social Adm Pharm* 2016;12(2):293-301 [FREE Full text] [doi: [10.1016/j.sapharm.2015.05.001](https://doi.org/10.1016/j.sapharm.2015.05.001)] [Medline: [26048710](https://pubmed.ncbi.nlm.nih.gov/26048710/)]
18. Coffin PO, Tracy M, Bucciarelli A, Ompad D, Vlahov D, Galea S. Identifying injection drug users at risk of nonfatal overdose. *Acad Emerg Med* 2007 Jul;14(7):616-623 [FREE Full text] [doi: [10.1197/j.aem.2007.04.005](https://doi.org/10.1197/j.aem.2007.04.005)] [Medline: [17554010](https://pubmed.ncbi.nlm.nih.gov/17554010/)]
19. Piper TM, Rudenstine S, Stancliff S, Sherman S, Nandi V, Clear A, et al. Overdose prevention for injection drug users: lessons learned from naloxone training and distribution programs in New York City. *Harm Reduct J* 2007 Jan 25;4:3 [FREE Full text] [doi: [10.1186/1477-7517-4-3](https://doi.org/10.1186/1477-7517-4-3)] [Medline: [17254345](https://pubmed.ncbi.nlm.nih.gov/17254345/)]
20. Somerville NJ, O'Donnell J, Gladden RM, Zibbell JE, Green TC, Younkin M, et al. Characteristics of fentanyl overdose - Massachusetts, 2014-2016. *MMWR Morb Mortal Wkly Rep* 2017 Apr 14;66(14):382-386 [FREE Full text] [doi: [10.15585/mmwr.mm6614a2](https://doi.org/10.15585/mmwr.mm6614a2)] [Medline: [28406883](https://pubmed.ncbi.nlm.nih.gov/28406883/)]
21. The Massachusetts Opioid Epidemic: An Issue of Substance. Massachusetts Taxpayers Foundation. 2018 Nov. URL: <https://rizema.org/wp-content/uploads/2018/11/mtf-opioid-report-final.pdf> [accessed 2021-05-04]
22. Pardo B, Taylor J, Caulkins JP, Kilmer B, Reuter P, Stein BD. Understanding America's Surge in Fentanyl and Other Synthetic Opioids. RAND Corporation. 2019. URL: [https://www.rand.org/pubs/research\\_briefs/RB10091.html](https://www.rand.org/pubs/research_briefs/RB10091.html) [accessed 2021-07-14]
23. Rev. URL: <https://www.rev.com/> [accessed 2021-06-21]
24. Glaser BG. The constant comparative method of qualitative analysis. *Grounded Theory Rev* 1965 Apr 1;12(4):436-445.
25. Bowles JM, Smith LR, Verdugo SR, Wagner KD, Davidson PJ. "Generally, you get 86'ed because you're a liability": an application of Integrated Threat Theory to frequently witnessed overdoses and social distancing responses. *Soc Sci Med* 2020 Sep;260:113190 [FREE Full text] [doi: [10.1016/j.socscimed.2020.113190](https://doi.org/10.1016/j.socscimed.2020.113190)] [Medline: [32673794](https://pubmed.ncbi.nlm.nih.gov/32673794/)]
26. Kahn LS, Wozniak M, Vest BM, Moore C. "Narcan encounters:" overdose and naloxone rescue experiences among people who use opioids. *Subst Abus* 2022;43(1):113-126. [doi: [10.1080/08897077.2020.1748165](https://doi.org/10.1080/08897077.2020.1748165)] [Medline: [32243234](https://pubmed.ncbi.nlm.nih.gov/32243234/)]
27. Urmanche AA, Beharie N, Harocopos A. Fentanyl preference among people who use opioids in New York City. *Drug Alcohol Depend* 2022 Aug 01;237:109519. [doi: [10.1016/j.drugalcdep.2022.109519](https://doi.org/10.1016/j.drugalcdep.2022.109519)] [Medline: [35714532](https://pubmed.ncbi.nlm.nih.gov/35714532/)]
28. McLean K, Monnat SM, Rigg K, Sterner 3rd GE, Verdery A. "You never know what you're getting": opioid users' perceptions of fentanyl in Southwest Pennsylvania. *Subst Use Misuse* 2019;54(6):955-966 [FREE Full text] [doi: [10.1080/10826084.2018.1552303](https://doi.org/10.1080/10826084.2018.1552303)] [Medline: [30676198](https://pubmed.ncbi.nlm.nih.gov/30676198/)]
29. Rowe C, Santos GM, Behar E, Coffin PO. Correlates of overdose risk perception among illicit opioid users. *Drug Alcohol Depend* 2016 Feb 01;159:234-239 [FREE Full text] [doi: [10.1016/j.drugalcdep.2015.12.018](https://doi.org/10.1016/j.drugalcdep.2015.12.018)] [Medline: [26754425](https://pubmed.ncbi.nlm.nih.gov/26754425/)]

30. Behar E, Chang JS, Countess K, Matheson T, Santos GM, Coffin P. Perceived causes of personal versus witnessed overdoses among people who inject opioids. *Subst Use Misuse* 2019;54(12):1929-1937 [[FREE Full text](#)] [doi: [10.1080/10826084.2019.1609988](https://doi.org/10.1080/10826084.2019.1609988)] [Medline: [31070106](https://pubmed.ncbi.nlm.nih.gov/31070106/)]
31. Goldman JE, Wayne KM, Periera KA, Krieger MS, Yedinak JL, Marshall BD. Perspectives on rapid fentanyl test strips as a harm reduction practice among young adults who use drugs: a qualitative study. *Harm Reduct J* 2019 Jan 08;16(1):3 [[FREE Full text](#)] [doi: [10.1186/s12954-018-0276-0](https://doi.org/10.1186/s12954-018-0276-0)] [Medline: [30621699](https://pubmed.ncbi.nlm.nih.gov/30621699/)]
32. Bennett AS, Freeman R, Des Jarlais DC, Aronson ID. Reasons people who use opioids do not accept or carry no-cost naloxone: qualitative interview study. *JMIR Form Res* 2020 Dec 23;4(12):e22411 [[FREE Full text](#)] [doi: [10.2196/22411](https://doi.org/10.2196/22411)] [Medline: [33355094](https://pubmed.ncbi.nlm.nih.gov/33355094/)]
33. Bennett AS, Scheidell J, Bowles JM, Khan M, Roth A, Hoff L, et al. Naloxone protection, social support, network characteristics, and overdose experiences among a cohort of people who use illicit opioids in New York City. *Harm Reduct J* 2022 Mar 04;19(1):20 [[FREE Full text](#)] [doi: [10.1186/s12954-022-00604-w](https://doi.org/10.1186/s12954-022-00604-w)] [Medline: [35246165](https://pubmed.ncbi.nlm.nih.gov/35246165/)]
34. Latimore AD, Bergstein RS. "Caught with a body" yet protected by law? Calling 911 for opioid overdose in the context of the Good Samaritan Law. *Int J Drug Policy* 2017 Dec;50:82-89. [doi: [10.1016/j.drugpo.2017.09.010](https://doi.org/10.1016/j.drugpo.2017.09.010)] [Medline: [29040841](https://pubmed.ncbi.nlm.nih.gov/29040841/)]
35. Wagner KD, Harding RW, Kelley R, Labus B, Verdugo SR, Copulsky E, et al. Post-overdose interventions triggered by calling 911: centering the perspectives of people who use drugs (PWUDs). *PLoS One* 2019 Oct 17;14(10):e0223823 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0223823](https://doi.org/10.1371/journal.pone.0223823)] [Medline: [31622401](https://pubmed.ncbi.nlm.nih.gov/31622401/)]
36. Davidson PJ, Ochoa KC, Hahn JA, Evans JL, Moss AR. Witnessing heroin-related overdoses: the experiences of young injectors in San Francisco. *Addiction* 2002 Dec;97(12):1511-1516. [doi: [10.1046/j.1360-0443.2002.00210.x](https://doi.org/10.1046/j.1360-0443.2002.00210.x)] [Medline: [12472634](https://pubmed.ncbi.nlm.nih.gov/12472634/)]
37. Tracy M, Piper TM, Ompad D, Bucciarelli A, Coffin PO, Vlahov D, et al. Circumstances of witnessed drug overdose in New York City: implications for intervention. *Drug Alcohol Depend* 2005 Aug 01;79(2):181-190. [doi: [10.1016/j.drugalcdep.2005.01.010](https://doi.org/10.1016/j.drugalcdep.2005.01.010)] [Medline: [16002027](https://pubmed.ncbi.nlm.nih.gov/16002027/)]
38. Durieux J, Curtis A, Mirka M, Jefferis E, Felix C, Essel B. An exploration of Narcan as a harm reduction strategy and user's attitudes toward law enforcement involvement in overdose cases. *Int J Environ Res Public Health* 2022 Mar 08;19(6):3149 [[FREE Full text](#)] [doi: [10.3390/ijerph19063149](https://doi.org/10.3390/ijerph19063149)] [Medline: [35328837](https://pubmed.ncbi.nlm.nih.gov/35328837/)]
39. Ataiants J, Mazzella S, Roth AM, Sell RL, Robinson LF, Lankenau SE. Overdose response among trained and untrained women with a history of illicit drug use: a mixed-methods examination. *Drugs (Abingdon Engl)* 2021;28(4):328-339 [[FREE Full text](#)] [doi: [10.1080/09687637.2020.1818691](https://doi.org/10.1080/09687637.2020.1818691)] [Medline: [34321719](https://pubmed.ncbi.nlm.nih.gov/34321719/)]
40. Rochester E, Graboyes M. Experiences of people who use drugs with naloxone administration: a qualitative study. *Drugs (Abingdon Engl)* 2022;29(1):54-61. [doi: [10.1080/09687637.2020.1855116](https://doi.org/10.1080/09687637.2020.1855116)]
41. van der Meulen E, Chu SK, Butler-McPhee J. "That's why people don't call 911": ending routine police attendance at drug overdoses. *Int J Drug Policy* 2021 Feb;88:103039. [doi: [10.1016/j.drugpo.2020.103039](https://doi.org/10.1016/j.drugpo.2020.103039)] [Medline: [33227637](https://pubmed.ncbi.nlm.nih.gov/33227637/)]
42. Opioid overdose basics: responding to an opioid overdose. National Harm Reduction Coalition. 2020 Sep 1. URL: <https://harmreduction.org/issues/overdose-prevention/overview/overdose-basics/responding-to-opioid-overdose/> [accessed 2022-06-09]
43. Morse JM. Determining sample size. *Qual Health Res* 2000 Jan 01;10(1):3-5. [doi: [10.1177/104973200129118183](https://doi.org/10.1177/104973200129118183)]
44. Latkin CA, Edwards C, Davey-Rothwell MA, Tobin KE. The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. *Addict Behav* 2017 Oct;73:133-136 [[FREE Full text](#)] [doi: [10.1016/j.addbeh.2017.05.005](https://doi.org/10.1016/j.addbeh.2017.05.005)] [Medline: [28511097](https://pubmed.ncbi.nlm.nih.gov/28511097/)]
45. Harris MT, Bagley SM, Maschke A, Schoenberger SF, Sampath S, Walley AY, et al. Competing risks of women and men who use fentanyl: "the number one thing I worry about would be my safety and number two would be overdose". *J Subst Abuse Treat* 2021 Jun;125:108313 [[FREE Full text](#)] [doi: [10.1016/j.jsat.2021.108313](https://doi.org/10.1016/j.jsat.2021.108313)] [Medline: [34016300](https://pubmed.ncbi.nlm.nih.gov/34016300/)]

---

## Abbreviations

**ADAPT-ITT:** assessment, decision, adaptation, production, topical experts, integration, training, and testing

**MOUD:** medications for opioid use disorder

**REBOOT:** repeated-dose behavioral intervention to reduce opioid overdose

**SKOOP:** Skills and Knowledge on Overdose Prevention

---

*Edited by A Mavragani; submitted 22.02.22; peer-reviewed by S Linder, J Simmons; comments to author 18.04.22; revised version received 13.05.22; accepted 15.07.22; published 07.09.22.*

*Please cite as:*

*McMahan VM, Arenander J, Matheson T, Lambert AM, Brennan S, Green TC, Walley AY, Coffin PO*

*“There’s No Heroin Around Anymore. It’s All Fentanyl.” Adaptation of an Opioid Overdose Prevention Counseling Approach to Address Fentanyl Overdose: Formative Study*

*JMIR Form Res 2022;6(9):e37483*

*URL: <https://formative.jmir.org/2022/9/e37483>*

*doi: [10.2196/37483](https://doi.org/10.2196/37483)*

*PMID: [36069781](https://pubmed.ncbi.nlm.nih.gov/36069781/)*

©Vanessa M McMahan, Justine Arenander, Tim Matheson, Audrey M Lambert, Sarah Brennan, Traci C Green, Alexander Y Walley, Phillip O Coffin. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 07.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Objective Monitoring of Facioscapulohumeral Dystrophy During Clinical Trials Using a Smartphone App and Wearables: Observational Study

Ghobad Maleki<sup>1,2\*</sup>, BSc; Ahnjili Zhuparris<sup>1,2\*</sup>, MSc; Ingrid Koopmans<sup>1</sup>, MSc; Robert J Doll<sup>1</sup>, PhD; Noline Voet<sup>3,4</sup>, MD, PhD; Adam Cohen<sup>1</sup>, MD, PhD; Emilie van Brummelen<sup>1</sup>, PhD; Geert Jan Groeneveld<sup>1,2</sup>, MD, PhD; Joris De Maeyer<sup>5</sup>, PhD

<sup>1</sup>Centre for Human Drug Research, Leiden, Netherlands

<sup>2</sup>Leiden University Medical Center, Leiden, Netherlands

<sup>3</sup>Radboud University Medical Center, Nijmegen, Netherlands

<sup>4</sup>Klimmendaal, Arnhem, Netherlands

<sup>5</sup>Facio Therapies, Leiden, Netherlands

\*these authors contributed equally

**Corresponding Author:**

Joris De Maeyer, PhD

Facio Therapies

Galileiweg 8

Leiden, 2333BD

Netherlands

Phone: 31 496109262

Email: [joris.demaeyer@facio-therapies.com](mailto:joris.demaeyer@facio-therapies.com)

## Abstract

**Background:** Facioscapulohumeral dystrophy (FSHD) is a progressive muscle dystrophy disorder leading to significant disability. Currently, FSHD symptom severity is assessed by clinical assessments such as the FSHD clinical score and the Timed Up-and-Go test. These assessments are limited in their ability to capture changes continuously and the full impact of the disease on patients' quality of life. Real-world data related to physical activity, sleep, and social behavior could potentially provide additional insight into the impact of the disease and might be useful in assessing treatment effects on aspects that are important contributors to the functioning and well-being of patients with FSHD.

**Objective:** This study investigated the feasibility of using smartphones and wearables to capture symptoms related to FSHD based on a continuous collection of multiple features, such as the number of steps, sleep, and app use. We also identified features that can be used to differentiate between patients with FSHD and non-FSHD controls.

**Methods:** In this exploratory noninterventional study, 58 participants (n=38, 66%, patients with FSHD and n=20, 34%, non-FSHD controls) were monitored using a smartphone monitoring app for 6 weeks. On the first and last day of the study period, clinicians assessed the participants' FSHD clinical score and Timed Up-and-Go test time. Participants installed the app on their Android smartphones, were given a smartwatch, and were instructed to measure their weight and blood pressure on a weekly basis using a scale and blood pressure monitor. The user experience and perceived burden of the app on participants' smartphones were assessed at 6 weeks using a questionnaire. With the data collected, we sought to identify the behavioral features that were most salient in distinguishing the 2 groups (patients with FSHD and non-FSHD controls) and the optimal time window to perform the classification.

**Results:** Overall, the participants stated that the app was well tolerated, but 67% (39/58) noticed a difference in battery life using all 6 weeks of data, we classified patients with FSHD and non-FSHD controls with 93% accuracy, 100% sensitivity, and 80% specificity. We found that the optimal time window for the classification is the first day of data collection and the first week of data collection, which yielded an accuracy, sensitivity, and specificity of 95.8%, 100%, and 94.4%, respectively. Features relating to smartphone acceleration, app use, location, physical activity, sleep, and call behavior were the most salient features for the classification.

**Conclusions:** Remotely monitored data collection allowed for the collection of daily activity data in patients with FSHD and non-FSHD controls for 6 weeks. We demonstrated the initial ability to detect differences in features in patients with FSHD and non-FSHD controls using smartphones and wearables, mainly based on data related to physical and social activity.

**Trial Registration:** ClinicalTrials.gov NCT04999735; <https://www.clinicaltrials.gov/ct2/show/NCT04999735>

(*JMIR Form Res* 2022;6(9):e31775) doi:[10.2196/31775](https://doi.org/10.2196/31775)

## KEYWORDS

facioscapulohumeral dystrophy; FSHD; smartphone; wearables; machine learning; classification; mobile phone

## Introduction

### Background

A recent Dutch population study on facioscapulohumeral dystrophy (FSHD) estimated that approximately 2000 people in the Netherlands and approximately 800,000 people worldwide are living with FSHD [1]. Often, early symptoms include difficulty whistling, smiling, and closing the eyelids while asleep. Weakening of the facial muscles is generally followed by scapular winging. This abnormal positioning of the shoulder bone impairs the movement of the shoulders and arms. Further weakening of the muscles is commonly observed in the upper arms and may progress to the hip girdle and lower legs in severe cases. Less visible symptoms of FSHD are chronic pain and fatigue [2]. In addition to the physical symptoms the diagnosis of FSHD comes with an emotional and social burden. The highly variable and unpredictable progression of the disease can have a strong impact on the quality of life [3,4]: 90% of the affected individuals have visible symptoms by the age of 20 years and 1 in 5 patients with FSHD eventually becomes wheelchair dependent [5].

No therapy is currently available that stops the progression of FSHD [6-9]. Patients thus have to rely on symptomatic treatment such as medical devices or surgical intervention [2]. The development of novel treatment options to delay or halt disease progression is currently under investigation. However, measuring the effect of such new treatments is complicated because disease progression is slow and no objective surrogate end points, predictive for clinical benefit, have been established. App-based technologies may help to more closely monitor FSHD symptom progression and evaluate potential treatment effects on a continuous basis.

Currently, FSHD symptom severity is assessed by clinical scoring of symptoms such as the FSHD clinical score or mobility performance tests such as the Timed Up-and-Go test (TUG) and Reachable Workspace assessment [10-12]. These clinical severity and functional scores have several drawbacks. Scores change very slowly over time [13], are assessed in a clinic at 1 specific moment, and do not cover the implications of the disease on social and physical activity during daily life. The progressive muscle weakness characterizing FSHD leads to massive changes in the way people live their lives, affecting how they get around, how they complete daily activities, and whether they can work or care for children. Therefore, assessing disease severity may be improved by not only measuring muscle function but also evaluating social and physical activity data. This study aimed to address this by first classifying disease

using a smartphone app and wearables to continuously remotely monitor features relating to biometric, physical, and social activities of patients with FSHD in comparison with those of non-FSHD controls. Subsequently, we performed a second analysis in which we aimed to assess disease severity. This analysis will be described in a different paper.

### Objectives

We investigated the feasibility of remotely monitoring multiple features such as step count, sleep, app use, and location tracking in patients with FSHD and non-FSHD controls. First, we evaluated the participants' tolerability of these devices. We then characterized the patients with FSHD and non-FSHD controls in terms of composites of social, physical, and biometric activities. We sought to (1) distinguish patients with FSHD from non-FSHD controls using a classification machine learning model and determine the minimum monitoring window needed to perform the classification and (2) identify which of the remotely monitored features were most salient in differentiating between the 2 groups.

## Methods

### Study Overview

We conducted a cross-sectional, noninterventional study in patients with FSHD and non-FSHD controls. A total of 58 participants (n=38, 66%, patients with genetically confirmed FSHD and n=20, 34%, non-FSHD controls) were included in this study at the Centre for Human Drug Research (CHDR) in Leiden, The Netherlands, between April 2019 and October 2019. Patients were recruited from The Netherlands and Belgium.

### Ethics Approval

This study was performed in compliance with International Council for Harmonisation Good Clinical Practice and approved by the Stichting Beoordeling Ethiek Biomedisch Onderzoek Medical Ethics Committee (Assen, The Netherlands; CCMO number NL69288.056.19) according to Wet medisch-wetenschappelijk onderzoek met mensen (Dutch law on medical-scientific research with humans).

### Patient Population

To represent the clinical FSHD spectrum based on symptom severity and age, up to 40 patients with FSHD (and also 20 control participants) were deemed sufficient. As this study was exploratory, sample size was not based on power calculations.

Eligible patients with FSHD were aged >16 years, had genetically confirmed FSHD (FSHD1 or FSHD2), were symptomatic as demonstrated by the FSHD clinical score of

>0, and had an Android phone that they used as their main phone or were willing to use one for the duration of the study period. Patients with any comorbidity, expected to affect the measurements, were excluded. Eligible control participants were included using the same inclusion and exclusion criteria that were used to recruit the patients, except they did not have a diagnosis or symptoms of FSHD.

## Data Collection

### Clinical Assessments

On the first and last days of the study period, the FSHD clinical score assessment was performed in the group consisting of patients with FSHD, whereas the TUG was performed in both groups. On day 42 in both groups the user experience was assessed and the perceived burden questionnaire ([Multimedia Appendix 1](#)) administered.

The FSHD clinical score is a standardized clinical score that quantifies muscle weakness by combining the functional evaluations of the 6 muscle groups affected in FSHD. The scale

is divided into 6 independent sections that assess the strength and the functionality of facial muscles, scapular girdle muscles, upper limb muscles, distal leg muscles, pelvic girdle muscles, and abdominal muscles [11]. The TUG assesses mobility and balance by measuring the time it takes for a participant to stand up from a seated position in a chair, walk 3 meters, turn around, walk back 3 meters, and sit down again [12]. The user experience and perceived burden questionnaire was developed by the CHDR to measure the impact of remote monitoring of apps on smartphone performance. The questions are based on the overall experience of CHDR with mobile apps.

### Remote Monitoring Platform

All participants were remotely monitored using the CHDR Monitoring Remotely (CHDR MORE) platform for 42 days. CHDR MORE is a highly customizable platform that allows remote monitoring of participants using smartphones and wearables. The infrastructure used includes an Android app to collect data from smartphone sensors and a connection to the Withings Health (Withings) web-based platform to collect wearable data. All collected features are described in [Table 1](#).

**Table 1.** Overview of all smartphone and wearable sensors used in this study and their respective extracted features.

Device and sensor	Features
<b>Smartphone</b>	
Accelerometer	Maximum magnitude of the acceleration: 98%
Apps	Number of times an app is opened; amount of time app is open in foreground
GPS	Total kilometers traveled per day; average kilometers traveled per trip; 95% maximum distance from home
Google Places	Number of unique places visited; time spent at each unique location
Calls	Number of outgoing, incoming, and missed calls; number of calls from known and unknown contacts
Microphone	Percentage of time a human voice is present
<b>Wearables (Withings)</b>	
Watch step count	Total step count; mean steps per minute; mean steps per hour; maximum steps per hour
Watch heart rate	Heart rate: 5%, 50%, and 95% ranges and SD of heart rate percentage of time spent in resting heart rate
Watch sleep	Awake as well as light and deep sleep duration (minutes); number of awake as well as light and deep sleep periods; time to fall asleep (minutes)
Watch physical activity	Soft, moderate, and hard activity duration
Blood pressure monitor	Systolic and diastolic blood pressure
Scale	Weight (kg); muscle mass (kg); bone mass (kg); body fat (%); body water (%)

### Smartwatch, Smart Scale, and Blood Pressure Monitor

In total, three commercially available Withings devices were used: (1) heart rate, step count, and sleep patterns were assessed by the Withings Steel HR smartwatch; (2) weight, BMI, and skeletal muscle mass were assessed by the Withings Body+ scale; and (3) systolic blood pressure and diastolic blood pressure were assessed by the Withings blood pressure monitor. Data from the Withings devices were collected on the phone using Bluetooth and sent to the Withings storage servers before being transferred to a CHDR server. Participants were instructed to wear the Withings Steel HR smartwatch continuously for the duration of the study, and they measured their weight and blood pressure themselves weekly using the Withings Body+ scale and Withings blood pressure monitor, respectively.

### Privacy

The data collection as part of this study may raise privacy and data safety concerns. Therefore, during development of the CHDR MORE app, we addressed these concerns by building in several measures to maximize privacy for all participants. First, all data sources such as SMS text messaging logs, phone calls, and microphone activation only report summative outcomes. These sources cannot send the content of messages or whole recordings to the CHDR servers. In addition, location data only report relative location instead of absolute GPS coordinates. Furthermore, all calculations such as human voice detection are performed on the Android phone itself and removed afterward and all personal data are coded and safely stored on certified CHDR servers.

## Statistical Analysis

### Data Preprocessing

The data preprocessing and analysis pipelines were developed using Python (version 3.6.0; Python Software Foundation). The Python library scikit-learn was used for the feature extraction and the development of the machine learning models [14]. All data were manually and visually inspected for missing data and outlier data. The identified outliers (eg, traveling 10,000 kilometers in a single day) were subsequently removed from the analysis. Missing or excluded data points were not imputed.

### Feature Extraction

As disease progression in FSHD is gradual, the FSHD clinical scores and TUG scores were expected to remain stable during the 6-week period. The daily features were therefore averaged across a defined time window (see the *Identification of Optimal Time Window* section for more information). Table 1 provides a simplified overview of the features that were extracted from the CHDR MORE app and Withings sensors.

### Feature Selection

Before fitting the classification models to the data set, features were excluded using manual and automated feature selection. The authors (AZ, RJD, AC, EvB, GJG, and JDM) of this paper manually excluded features based on the degree of missing data and the clinical relevance of the feature (eg, time spent on home and house apps were deemed clinically irrelevant). For the automated feature selection, variance inflation factor and stepwise regression were used to exclude multi-collinear features or features that did not provide additive information, respectively.

### Classification Models

We used 4 categories of data sets for the classification of patients with FSHD and non-FSHD controls. These categories include the composite data (all features), social data (smartphone features relating to social location, social and communication app use, and phone calls), physical activity data (smartwatch features), and biometric data (scale and blood pressure monitor features). We compared the performance of the logistic regression, random forest, and support vector machine classification models (Multimedia Appendix 2 [15-22]). The performance of these classification models was evaluated by the accuracy, sensitivity, specificity, and Matthews correlation coefficient (MCC). A grid search was performed to find the optimal hyperparameters (the parameters that determine the model's structure) that would yield the highest sensitivity and

specificity for each model. Furthermore, we performed a 5-fold stratified cross-validation. Cross-validation is a resampling method used to evaluate the prediction performance of the classification models. The data were divided into 5 equal subsets, with the same FSHD-to-non-FSHD ratio within each subset; the model was trained on 4 (80%) partitions of the data and tested on 1 (20%) partition. This procedure was repeated 5 times, with each partition serving as a test set once. The performance of each model validation was then averaged.

### Identification of Optimal Time Window

In total, 6 weeks of data were collected for this study. As continuous and periodic data collection for long periods of time can be expensive and increase the risk of data loss, we investigated the minimum time window needed for reliable classification. First, we used an incrementally increasing time window to train the classification model, starting from day 1 and adding 1 day until we included all 42 days of data. We examined which time window would yield the highest overall accuracy, sensitivity, and specificity. We compared the performances of 3 classification algorithms (least absolute shrinkage and selection operator [LASSO]-penalized logistic regression, random forest, and support vector machine) to classify patients with FSHD and non-FSHD controls using the incremental time windows. Second, we used the optimal time window to train the classification model and evaluated how stable the classification performance would be for the remaining 5 weeks of data. Here, we evaluated the stability of the algorithm based on the generalization error of the trained classification model [23].

## Results

### Data Collected

In total, 58 participants (n=38, 66%, patients with FSHD and n=20, 34%, non-FSHD controls) participated in the study. We did not meet our goal of 40 patients because of difficulties in recruiting patients in an acceptable time span.

The female-to-male ratio was the same in both populations; however, the median age of the control participants without FSHD was lower than that of their counterparts with FSHD. Table 2 illustrates the demographic and disease characteristics of the participants enrolled in this study. The FSHD clinical scores and TUG scores remained relatively stable during the 6-week period (with a maximum intraparticipant change of 1 point for the FSHD score and 0.63 seconds for the TUG score).

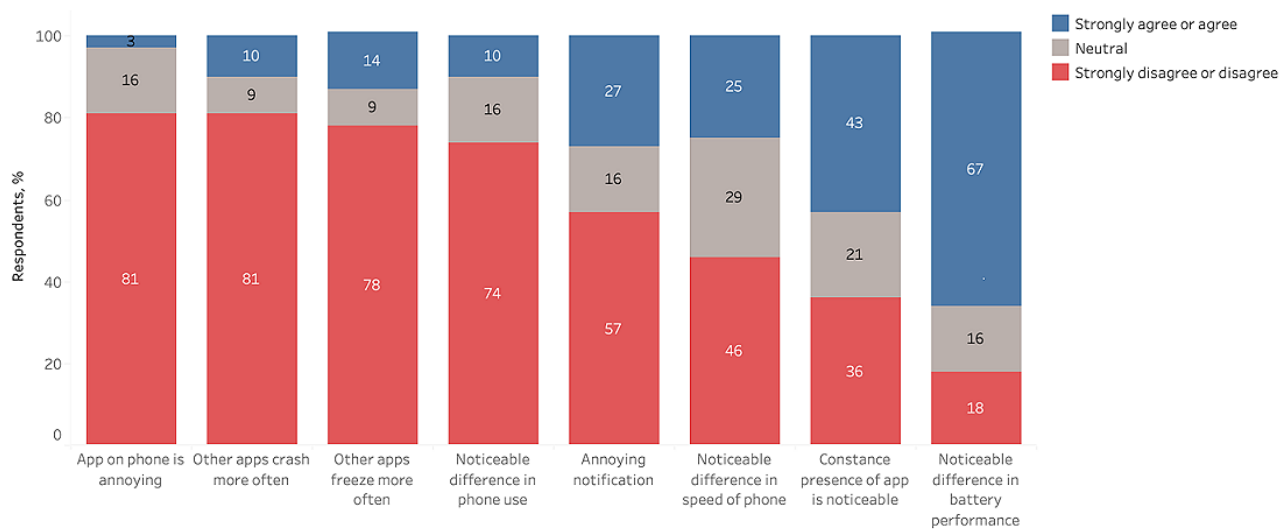
**Table 2.** Demographics of patients with facioscapulohumeral dystrophy (FSHD) and controls without FSHD (N=58).

Demographics	Patients with FSHD	Non-FSHD controls
<b>Sex, n (%)</b>		
Female	23 (61)	11 (55)
Male	15 (39)	9 (45)
Age (years), mean (SD; range)	45 (14.5; 18-64)	33 (12; 23-69)
Weight (kg), mean (SD; range)	80 (16; 52-130)	78 (18; 56-129)
BMI (kg/m <sup>2</sup> ), mean (SD; range)	26 (4; 20-44)	25 (5; 19-35)
FSHD clinical score, mean (SD; range)	5 (3; 1-13)	0 (0; 0-0)
Timed Up-and-Go test (seconds), mean (SD; range)	8.8 (35; 5-15.81)	7.8 (1.55; 6-12.09)

### Perceived Burden and Data Loss

As shown in [Figure 1](#), overall, 3% (2/58) of the participants found the app on their phone to be annoying. Furthermore, 67% (39/58) of the participants agreed that there was a noticeable difference in battery life, 43% (25/58) agreed that the constant presence of the app was noticeable on their smartphone, 28% (16/58) rated the constant visible notification as annoying, and 26% (15/58) of the participants noted a difference in the speed of their smartphone.

Data completeness is defined as having incoming data for each day of the clinical trial, except for the blood pressure and scale data, for which completeness is defined as having incoming data each week. As phone and SMS text messaging data are activity triggered and are aperiodic, it is not possible to know whether data were missing. [Table 3](#) provides an overview of data completeness for the CHDR MORE app, Withings watch, Withings scale, and Withings blood pressure monitor and their respective sensors.

**Figure 1.** Feasibility and perceived burden of remote monitoring in patients with facioscapulohumeral dystrophy using smartphone-based technologies.

**Table 3.** Overview of data completeness. The data completeness shows what percentage of data was collected among the participants during the 42 days of the study; hence, in total, there should be 2436 daily instances and 232 weekly instances.

Sensor	Feature	Overall data completion			
		Patients with FSHD <sup>a</sup>		Controls without FSHD	
		n (%)	N	n (%)	N
Microphone (smartphone)	Voice activation	1181 (74)	1596	688 (81.9)	840
Accelerometer (smartphone)	Phone acceleration	1260 (78.95)	1596	656 (78)	840
Google Places (smartphone)	Places	1109 (69.49)	1596	616 (73.33)	840
GPS (smartphone)	Relative location	1373 (86.03)	1596	785 (93.45)	840
App use (smartphone)	Use event aggregate	1404 (87.97)	1596	779 (92.74)	840
Withings blood pressure monitor	Blood pressure and heart rate	1452 (91.15)	1596	630 (75)	840
Withings scale	Body composition	173 (75.88)	228	88 (73.33)	120
Withings scale	Weight	205 (89.91)	228	108 (90)	120
Withings watch	Activity duration	1505 (94.3)	1596	744 (88.57)	840
Withings watch	Heart rate	1181 (74)	1596	588 (70)	840
Withings watch	Step count	1491 (93.42)	1596	708 (84.29)	840
Withings watch	Sleep summary	1408 (88.22)	1596	685 (81.55)	840

<sup>a</sup>FSHD: facioscapulohumeral dystrophy.

## Feature Selection

Several features were manually excluded before modeling. Because of the number of participants missing body composition data, we excluded all the body composition data with the exception of weight. Furthermore, we excluded SMS text message use features and app categories that were only used by only 5% (3/58) of the participants.

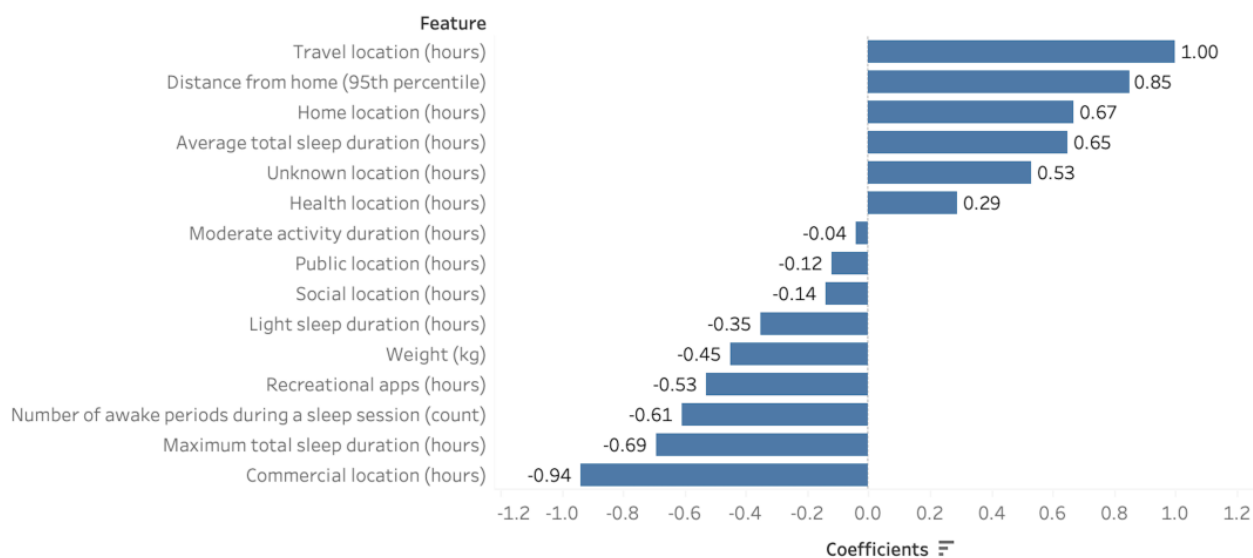
## Identification of Optimal Time Window and Classification Performance

Using all 6 weeks of data, the optimal classification model (LASSO-penalized logistic regression) achieved 93% accuracy, 100% sensitivity, 80% specificity, and 85% MCC. This classification model identified 15 features that were relevant for differentiating between patients with FSHD and non-FSHD controls. Specifically, features such as app use, weight, location, physical activity, and sleep were important for differentiating between the 2 populations (Figure 2). Table 4 shows the predictive features and their positive or negative associations with the classification label. The predictive features indicate that the participants in the group consisting of patients with FSHD were less likely to engage in moderate physical activity and spend time on recreational apps such as entertainment apps, music and audio apps, video players and editing apps, and games. The predictive features also showed that the participants in the group consisting of patients with FSHD were more likely to spend more time at home and health locations than their non-FSHD counterparts. Table 5 provides a summary of the number of selected features and the respective performance metric for each of the data sets fitted to the 6-week

LASSO-penalized logistic regression model. The table illustrates that the composite data set model outperformed the models fitted to the social, physical activity, and biometric data sets. The MCC is used to select the best model because it corrects for class imbalances. The scores of the individual data sets are included to give an overview of their performance on their own. The MCC values of the social activity, physical activity, and biometric logistic regression models were 52%, 38%, and -21%, respectively.

As for identifying the optimal time window for accurately classifying the patients with FSHD and non-FSHD controls, we found that training the random forest on the data collected on the first day and the data collected during the first week yielded an accuracy, sensitivity, specificity, and MCC of 95.8%, 100%, 94.4%, and 93.8% (Figure 3). This approach outperformed the classification models that were trained on all 6 weeks of data. We also trained classification models on the first week's data and fitted the data from subsequent weeks to assess the stability of the classification performance over time (Figure 4). We found that the random forest achieved the best overall performance, with a mean accuracy, sensitivity, specificity, and MCC of 95% (SD 0.9%), 97.6% (SD 3.6%), 94.1% (SD 0.9%), and 93.6% (SD 0.1%), respectively. Figure 5 provides a Shapley additive explanations plot that illustrates the magnitude and direction of the effect of a feature on a prediction. Of the 20 selected features, the top 5 (25%) most important features for the classification were mean kilometers traveled, 95% maximum distance from home, total kilometers traveled, 95% highest heart rate, and intense activity duration. For each of these features, the participants in the group consisting of patients with FSHD had lower scores than the non-FSHD controls.

**Figure 2.** Selected features for classifying patients with facioscapulohumeral dystrophy and those without facioscapulohumeral dystrophy based on the composite data set using all 6 weeks of data and the least absolute shrinkage and selection operator–penalized logistic regression model. Unstandardized estimated coefficients indicate the direction of the association between the feature and the classification label.



**Table 4.** Selected features for classifying patients with facioscapulohumeral dystrophy and controls without facioscapulohumeral dystrophy based on the complete 6-week composite data set. Unstandardized estimated coefficients indicate the direction of the association between the feature and the classification label.

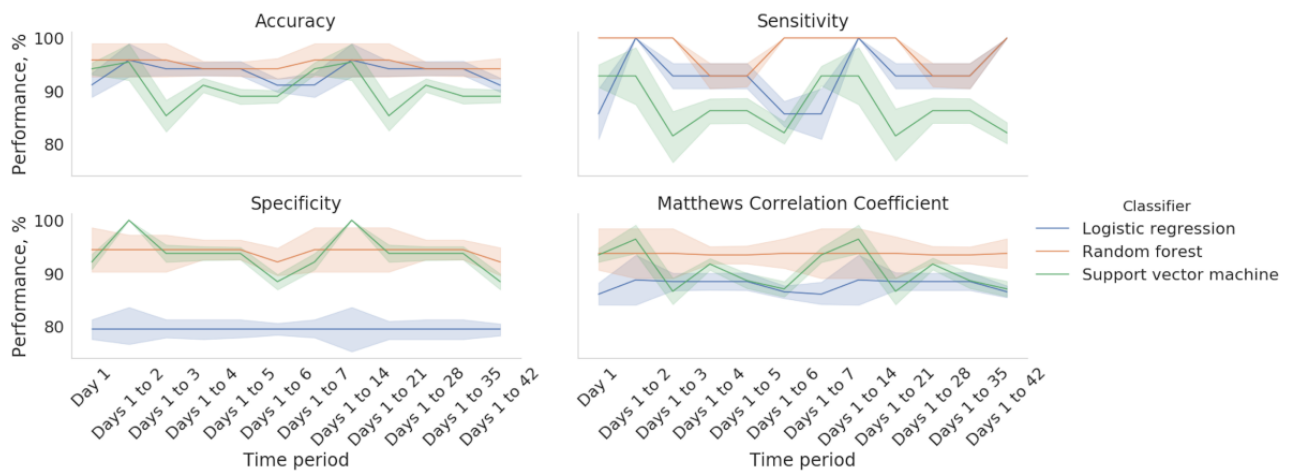
Feature category and feature	Unstandardized estimated coefficient
<b>Activity</b>	
Moderate activity duration	-0.04
<b>App</b>	
Time spent on recreational apps	-0.53
<b>Body</b>	
Weight (kg)	-0.45
<b>Location</b>	
Distance from home: 95%	0.85
<b>Time spent at location</b>	
Travel location	1.00
Home location	0.67
Unknown location	0.53
Health location	0.29
Public location	-0.12
Social location	-0.14
Commercial location	-0.94
<b>Sleep</b>	
Average total sleep duration	0.65
Light sleep duration	-0.35
Number of awake periods during a sleep session	-0.61
Maximum total sleep duration	-0.69

**Table 5.** Summary of number of selected features and the respective performance metric for each of the data sets used to classify the patients with facioscapulohumeral dystrophy from the controls without facioscapulohumeral dystrophy.

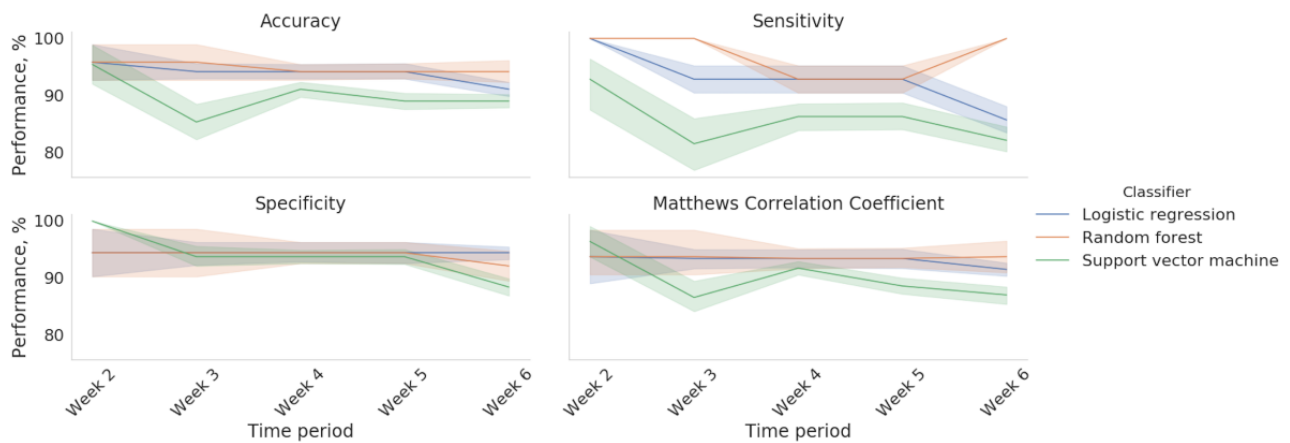
Data set	Number of selected features	Accuracy (%)	Sensitivity (%)	Specificity (%)	MCC <sup>a</sup> (%)
Composite	15	93	100	80	85
Biometric	5	57	89	0	-21
Social	10	79	90	60	52
Physical activity	13	71	78	60	38

<sup>a</sup>MCC: Matthews correlation coefficient.

**Figure 3.** Performance of the incremental classification predictions for 3 classifiers (logistic regression, random forest, and support vector machine). The x-axis shows the time window for training the classification models starting from day 1 to day 42. The error bands represent the SD of the classification performance for the 5-fold cross-validation.

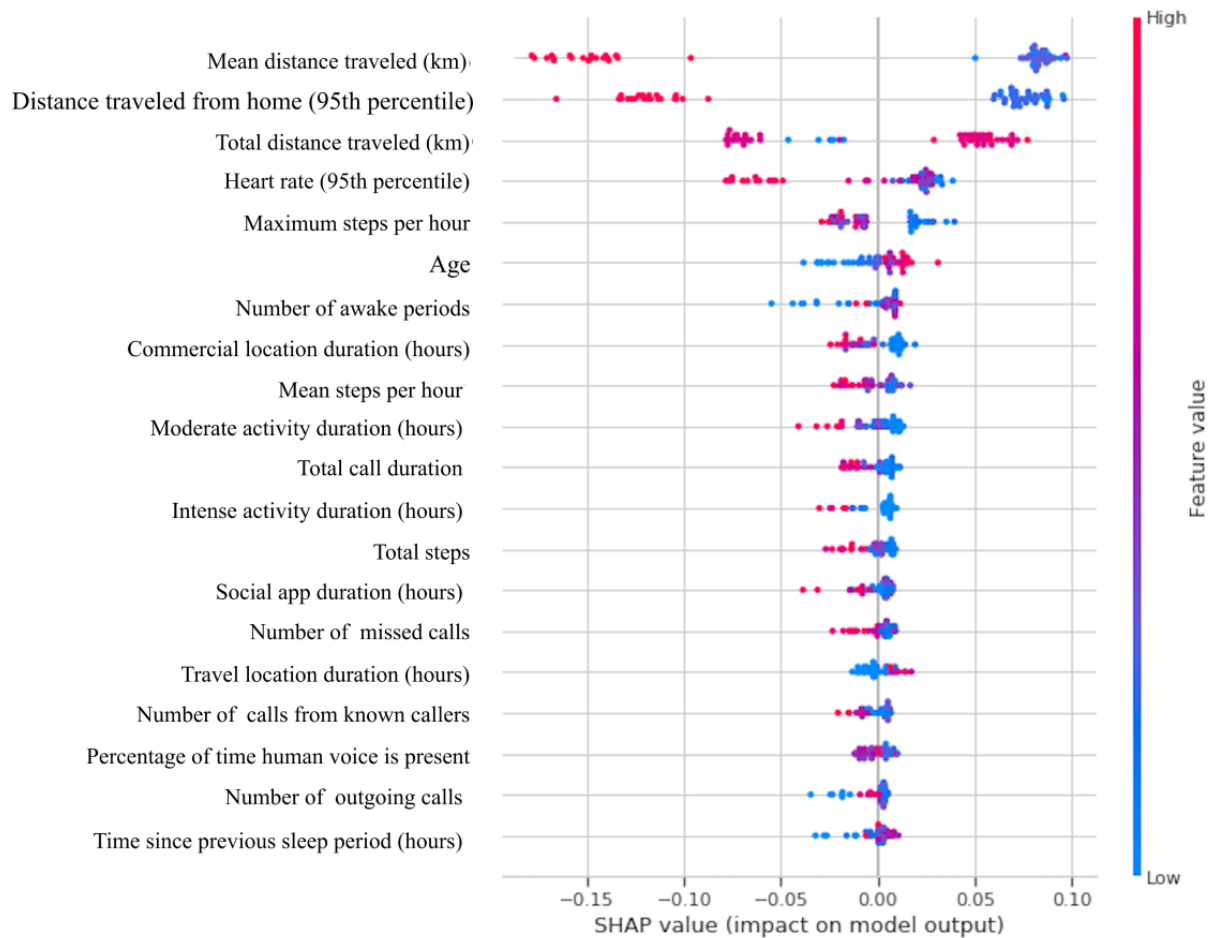


**Figure 4.** Performance of 3 classifiers (logistic regression, random forest, and support vector machine) trained on the week 1 data and used to predict the classification diagnosis of the subsequent weeks of data. The error bands represent the SD of the classification performance for the 5-fold cross-validation.





**Figure 5.** Shapley additive explanations (SHAP) summary plot based on a random forest classifier that was trained on the week 1 data. The x-axis shows the feature importance, where features are ranked in descending order. The y-axis shows the SHAP value that illustrates the direction of the association between the feature and facioscapulohumeral dystrophy severity. The color scheme reflects the probability of a participant being classified as a patient with facioscapulohumeral dystrophy.



## Discussion

### Principal Findings

We investigated the feasibility of monitoring and characterizing the physical, social, and biometric features of patients with FSHD and non-FSHD controls using remote monitoring technologies. The use of the remote monitoring platform was well tolerated by all participants. Next, we found that a minimum of 1 day of data and a maximum of 1 week of data can be used to reliably classify the 2 populations. In fact, an FSHD classification model trained on data from a shorter time window outperformed a classification model trained on data from the entire 6-week period. Furthermore, we illustrated that a classification model trained on the first week's data yielded stable and reliable classification predictions across the remaining 5-week period.

Most (37/58, 64%) of the participants tolerated the CHDR MORE app constantly running on their smartphone (Figure 1). Of the 58 participants, only 2 (3%) stated that the app was annoying. However, the results show that some of the participants agreed that there was a noticeable difference in smartphone speed performance (14/58, 25%), stability (8/58, 14%), and overall battery life (39/58, 67%). Therefore, the presence of the app was noticeable for some (25/58, 43%) of

the participants. The decrease in smartphone performance (ie, speed, stability, and battery performance) was likely due to the continuous sampling of the sensors. As this was the first study in this specific patient group with this platform, all smartphone sensors were frequently sampled to capture all possible features. With the collected data in this study, we identified the features that are useful in differentiating between patients with FSHD and non-FSHD controls. In future studies, noncontributing raw data such as data from the accelerometer and gyroscope (both sampled at 5 Hz) can be turned off to reduce the burden on the battery performance and overall user experience. We do not know for certain whether, and how, the noticeability of the app affects participants' behavior. Of the 58 participants, 6 (10%) stated that they noticed a change in smartphone use for themselves, which may mean that they changed their behavior. Therefore, participants will know that they are participating in a study and that they are being constantly monitored even if the app is perfectly optimized. As a result, some sort of change in behavior can be expected.

As for the user experience and perceived burden questionnaire, we designed a questionnaire based on our own experiences with smartphone use and the predicted effects of the CHDR MORE app on smartphones. This questionnaire was not validated in any other study. At the time of designing the study, there were no validated and published smartphone app questionnaires that

would fit our purpose. For example, the mHealth App Usability Questionnaire [24] focuses more on *active* smartphone apps, where there is interaction between the app and the participants. The CHDR MORE app is a *passive* app, requiring almost no interaction between the app and the user. Therefore, the questions should be more focused on the indirect effects of the app, such as more frequent crashes in other apps, subjective loss of snappiness of the operating system, or issues with battery performance. Although our questionnaire is not validated, it was considered the best way to accurately capture the perceived impact of the CHDR MORE app on smartphone use.

Feature selection is one of the most important processes for building a classification model. The inclusion of irrelevant features can confound the interpretability of the model because potentially predictive features would be excluded and therefore seem to be irrelevant. For example, because the patients with FSHD had more text-related activity than the non-FSHD controls, the SMS text messaging features were selected as important classification features. Given that the SMS text messaging features were not deemed clinically relevant because only 55% (21/38) of the patients with FSHD and 50% (10/20) of the non-FSHD controls actively sent outgoing SMS text messages and the majority of the SMS text messages were exchanged with unknown contacts, we excluded the SMS text messages as a feature. As a result, features that were initially not selected by the model for inclusion, such as sleep, were now deemed important features. The SMS text messaging features masked the relevance of other potentially predictive features. The features that researchers manually choose to include or exclude will influence the interpretability and stability of the model. It should be noted that although SMS text messaging features were excluded, features regarding instant messaging app use were included.

Our classification models allowed for the identification of a stable set of features that were distinctive of FSHD symptomology. We believe that identifying which remotely monitored features are relevant to FSHD can be a first step toward continuous monitoring of symptom severity and disease progression. For example, our classification model identified sleep as a relevant feature for classifying patients with FSHD. Other studies have found that patients with FSHD typically experience sleep anomalies because of anxiety, respiratory muscle dysfunction, and pain [25-27]. This illustrates that the CHDR MORE platform is sensitive enough to detect and monitor sleep anomalies among individuals with FSHD outside of the clinic. Furthermore, location-related features were relevant for differentiating between the 2 populations. In this study, the patients with FSHD spent more time at home, in areas with public transportation, or at health locations than the healthy participants. Patients with FSHD face a range of physical challenges because of the functional deterioration in the affected muscular regions. Consequently, patients with FSHD may become more home bound and more reliant on public transportation for travel, as well as require more visits to their physicians. In conclusion, the CHDR MORE platform provides data that can be used to show differences in the daily lives of patients with FSHD and controls without FSHD.

We demonstrated that there is a trade-off among the classification accuracy, the number of sensor measurements, and the duration of the monitoring period. Previous studies have demonstrated that using data from multiple sensors improves the detection of mental and physical health status compared with using data from a single sensor [28-30]. We illustrated that social activity, physical activity, and biometric data alone are insufficient for the accurate classification of FSHD. Rather, the inclusion of data from the smartphone, smartwatch, and scale improves the performance of the FSHD classification algorithm. Although the modeling of multi-sensor data can be advantageous, it can lead to several practical limitations. The inclusion of more features can increase the model's complexity and thus limit the model's explainability. Furthermore, the inclusion of more sensors and a longer monitoring period can be more expensive, potentially limit the number of participants enrolled in a study, and increase the risk of data loss. Future studies will need to weigh the advantages and disadvantages of integrating smartphones, smartwatches, scales, and monitoring period into their remotely monitored FSHD clinical trials.

Despite the good performance of our model, this study includes some limitations. The patients with FSHD and non-FSHD controls were comparable except for the age demographic. The median age of the non-FSHD controls was approximately 13 years less than that of the patients with FSHD. Generally, the older the person, the less they tend to use their smartphone and, in particular, the less they tend to use communication and social apps [31]. When characterizing patients with FSHD and non-FSHD controls based on active smartphone use, the model may be biased because of the difference in age. However, as seen in the results, only 1 feature of active smartphone use—time spent on recreational apps—was included in the final model for the characterization of patients with FSHD, which may limit the impact of this difference. The other features used in the composite model consist of either physical activity features collected passively from the smartphone or biometric data collected from the Withings devices. Therefore, we believe that the impact of these *contaminated* data on the performance of our model is low.

The objective of our study was to capture continuous sensor data. However, these data can only be considered reliable when participants carry their smartphone and have it turned on all the time. During this study, all participants were instructed to do so. However, data captured when the participant was not carrying their smartphone could not be distinguished from data captured when the participant was carrying the smartphone. Therefore, all instances in which the smartphone is not carried or turned on result in unrepresentative data. These data get mixed in the *real* data because these moments cannot be filtered out of the data with full certainty, resulting in unreliable data. Of note, there is no easy solution to this problem. It would be difficult to continuously check whether the participants are carrying their smartphone using the built-in sensors. However, adherence to this requirement is an important aspect in remote data collection, emphasizing the need for clear instructions on this adherence aspect to participants during training sessions before study start.

The level of data loss from the Withings scale indicates that improvement is needed to gather reliable scale data (Table 3). Data loss occurred for both the patients with FSHD and the non-FSHD controls, indicating that the loss of data was unlikely related to any of the FSHD symptoms. Although clear instructions were given at the beginning of the study and all participants received a manual with the same instructions, we believe that the data loss was caused by improper use of the scale by the participants. The weight measurement consisted of two parts: (1) measurement of weight and (2) measurement of body composition. Weight was determined first, followed by a blinking notification on the display during the measurement of body composition. This might have given the impression to the participant that the measurement had been completed, causing them to interrupt the second part of the measurement, resulting in an incomplete measurement. For future studies, we recommend incorporating a live training at the beginning of the study on the correct use of the scale.

Efficient clinical testing of any FSHD intervention or of any drug targeted at improving function of patients with FSHD or delaying disease progression requires the availability of clinical biomarkers that ideally change relatively rapidly over time; correlate with, and allow for, prediction of progression of the

existing clinical severity and functional scores; and allow for identification of fast progressors. Using data collected in a home setting might provide a more comprehensive picture of the evolution of a patient's overall condition over time. This study is a first step in the development and validation process of using data collected by a specific remote monitoring platform for use in patients with FSHD. The features described in this paper may be useful in further evaluating the impact of the disease and monitoring disease progression in patients with FSHD in the future [13]. More extensive data from longitudinal studies are needed to further define how social, physical, and biometric data collected remotely can be used to monitor symptoms.

## Conclusions

To conclude, this study illustrates that the collection of smartphone data and wearable data is acceptable to patients with FSHD and non-FSHD controls and can be used to differentiate between the 2 populations. We showed that remotely monitored end points can capture behavioral differences between patients and controls. Further longitudinal studies are warranted to study the potential of using a remote monitoring system for detecting FSHD symptom severity and possible drug effects.

---

## Acknowledgments

The authors thank all participants who participated in this study. The authors would like to thank Otto Postma for his contributions in designing this study and his input in the analysis. The authors would also like to thank Spierziekten Nederland and Spierziekten Vlaanderen for their help in the recruitment of patients. The study was funded by Facio Therapies, Leiden, The Netherlands.

---

## Data Availability

The data that support the findings of this study are available from Facio Therapies, but restrictions apply to the availability of these data. However, data are available from the authors upon reasonable request and with permission of Facio Therapies.

---

## Authors' Contributions

GM conceptualized the study, supported data collection, and wrote the manuscript. AZ conceptualized the study, performed the analysis, and wrote the manuscript. IK conceptualized the study, collected data, and supported the analysis. RJD conceptualized the study, supervised and provided input for the analysis, and reviewed the manuscript. GJG and EvB conceptualized the study, supervised the clinical part of the study, and reviewed the manuscript. AC, NV, and JDM conceptualized the study, provided input for the analysis, and reviewed the manuscript.

---

## Conflicts of Interest

At the time of the study, AC worked at the Centre for Human Drug Research as well as an unpaid member of the scientific advisory board of Facio Therapies. JDM is the Chief Executive Officer of Facio Therapies.

---

### Multimedia Appendix 1

Perceived burden questionnaire.

[DOCX File, 16 KB - [formative\\_v6i9e31775\\_app1.docx](#) ]

---

### Multimedia Appendix 2

Additional background information of investigated classifiers.

[DOCX File, 17 KB - [formative\\_v6i9e31775\\_app2.docx](#) ]

---

## References

1. Deenen JC, Arnts H, van der Maarel SM, Padberg GW, Verschuuren JJ, Bakker E, et al. Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* 2014 Aug 13;83(12):1056-1059. [doi: [10.1212/wnl.0000000000000797](https://doi.org/10.1212/wnl.0000000000000797)]
2. Voet N, Bleijenberg G, Hendriks J, de Groot I, Padberg G, van Engelen B, et al. Both aerobic exercise and cognitive-behavioral therapy reduce chronic fatigue in FSHD: an RCT. *Neurology* 2014 Oct 22;83(21):1914-1922. [doi: [10.1212/wnl.0000000000001008](https://doi.org/10.1212/wnl.0000000000001008)]
3. Padua L, Aprile I, Frusciante R, Iannaccone E, Rossi M, Renna R, et al. Quality of life and pain in patients with facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2009 Aug 16;40(2):200-205. [doi: [10.1002/mus.21308](https://doi.org/10.1002/mus.21308)] [Medline: [19609906](https://pubmed.ncbi.nlm.nih.gov/19609906/)]
4. Tawil R, McDermott MP, Mendell JR, Kissel J, Griggs RC. Facioscapulohumeral muscular dystrophy (FSHD): design of natural history study and results of baseline testing. FSH-DY Group. *Neurology* 1994 Mar 01;44(3 Pt 1):442-446. [doi: [10.1212/wnl.44.3\\_part\\_1.442](https://doi.org/10.1212/wnl.44.3_part_1.442)] [Medline: [8145913](https://pubmed.ncbi.nlm.nih.gov/8145913/)]
5. Wang LH, Tawil R. Facioscapulohumeral dystrophy. *Curr Neurol Neurosci Rep* 2016 Jul 23;16(7):66. [doi: [10.1007/s11910-016-0667-0](https://doi.org/10.1007/s11910-016-0667-0)] [Medline: [27215221](https://pubmed.ncbi.nlm.nih.gov/27215221/)]
6. Kissel JT, McDermott MP, Mendell JR, King WM, Pandya S, Griggs RC, FSH-DY Group. Randomized, double-blind, placebo-controlled trial of albuterol in facioscapulohumeral dystrophy. *Neurology* 2001 Oct 23;57(8):1434-1440. [doi: [10.1212/wnl.57.8.1434](https://doi.org/10.1212/wnl.57.8.1434)] [Medline: [11673585](https://pubmed.ncbi.nlm.nih.gov/11673585/)]
7. Kley R, Tamopolsky M, Vorgerd M. Creatine for treating muscle disorders. *Cochrane Database Systematic Rev* 2004(2):CD004760. [doi: [10.1002/14651858.CD004760](https://doi.org/10.1002/14651858.CD004760)]
8. Tawil R, McDermott MP, Pandya S, King W, Kissel J, Mendell JR, et al. A pilot trial of prednisone in facioscapulohumeral muscular dystrophy. FSH-DY Group. *Neurology* 1997 Jan 01;48(1):46-49. [doi: [10.1212/wnl.48.1.46](https://doi.org/10.1212/wnl.48.1.46)] [Medline: [9008492](https://pubmed.ncbi.nlm.nih.gov/9008492/)]
9. Le Gall L, Sidlauskaitė E, Mariot V, Dumonceaux J. Therapeutic strategies targeting DUX4 in FSHD. *J Clin Med* 2020 Sep 07;9(9):2886 [FREE Full text] [doi: [10.3390/jcm9092886](https://doi.org/10.3390/jcm9092886)] [Medline: [32906621](https://pubmed.ncbi.nlm.nih.gov/32906621/)]
10. Validation of the upper extremity reachable work space (RWS) as a clinical outcome assessment (COA) for FSHD clinical trials. Muscular Dystrophy Association. URL: <https://www.mdaconference.org/node/918> [accessed 2021-02-20]
11. Lamperti C, Fabbri G, Vercelli L, D'Amico R, Frusciante R, Bonifazi E, et al. A standardized clinical evaluation of patients affected by facioscapulohumeral muscular dystrophy: the FSHD clinical score. *Muscle Nerve* 2010 Aug 18;42(2):213-217. [doi: [10.1002/mus.21671](https://doi.org/10.1002/mus.21671)] [Medline: [20544930](https://pubmed.ncbi.nlm.nih.gov/20544930/)]
12. Podsiadlo D, Richardson S. *J Am Geriatr Soc* 1991 Feb;39(2):142-148. [doi: [10.1111/j.1532-5415.1991.tb01616.x](https://doi.org/10.1111/j.1532-5415.1991.tb01616.x)] [Medline: [1991946](https://pubmed.ncbi.nlm.nih.gov/1991946/)]
13. Kruizinga MD, Stuurman FE, Exadaktylos V, Doll RJ, Stephenson DT, Groeneveld GJ, et al. Development of novel, value-based, digital endpoints for clinical trials: a structured approach toward fit-for-purpose validation. *Pharmacol Rev* 2020 Oct 21;72(4):899-909. [doi: [10.1124/pr.120.000028](https://doi.org/10.1124/pr.120.000028)] [Medline: [32958524](https://pubmed.ncbi.nlm.nih.gov/32958524/)]
14. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825-2830 [FREE Full text] [doi: [10.5555/1953048.2078195](https://doi.org/10.5555/1953048.2078195)]
15. Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. *JAMA* 2016 Aug 02;316(5):533-534. [doi: [10.1001/jama.2016.7653](https://doi.org/10.1001/jama.2016.7653)] [Medline: [27483067](https://pubmed.ncbi.nlm.nih.gov/27483067/)]
16. Czepiel SA. Regression models. In: *Maximum Likelihood Estimation for Sample Surveys*. Milton Park, Abingdon-on-Thames, Oxfordshire United Kingdom: Taylor & Francis; 2012.
17. Tibshirani R. Regression shrinkage and selection via the lasso. *J Royal Stat Soc: series B (Methodological)* 2018 Dec 05;58(1):267-288. [doi: [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x)]
18. Ho T. Proceedings of 3rd International Conference on Document Analysis and Recognition. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. 1995 Presented at: Proceedings of 3rd International Conference on Document Analysis and Recognition; Aug 14-16, 1995; Montreal, QC, Canada URL: <https://ieeexplore.ieee.org/document/598929/citations#citations> [doi: [10.1109/icdar.1995.598929](https://doi.org/10.1109/icdar.1995.598929)]
19. Muchlinski D, Siroky D, He J, Kocher M. Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. *Polit anal* 2017 Jan 04;24(1):87-103. [doi: [10.1093/pan/mpv024](https://doi.org/10.1093/pan/mpv024)]
20. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995 Sep;20(3):273-297. [doi: [10.1007/bf00994018](https://doi.org/10.1007/bf00994018)]
21. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, Second Edition. Cham: Springer; 2009.
22. Suykens J, Vandewalle J. Least squares support vector machine classifiers. *Neural Processing Letters* 1999;9(3):293-300. [doi: [10.1023/A:1018628609742](https://doi.org/10.1023/A:1018628609742)]
23. Bousquet O, Elisseeff A. Stability and generalization. *J Mach Learn Res* 2002;2(3):499-526 [FREE Full text]
24. 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 [FREE Full text] [doi: [10.2196/11500](https://doi.org/10.2196/11500)] [Medline: [30973342](https://pubmed.ncbi.nlm.nih.gov/30973342/)]
25. Leclair-Visonneau L, Magot A, Tremblay A, Bruneau X, Pereon Y. P.16.5 Anxiety is responsible for altered sleep quality in Facio-Scapulo-Humeral Muscular Dystrophy (FSHD). *Neuromuscular Disorders* 2013 Oct;23(9-10):823-824. [doi: [10.1016/j.nmd.2013.06.642](https://doi.org/10.1016/j.nmd.2013.06.642)]

26. Della Marca G, Frusciante R, Vollono C, Iannaccone E, Dittoni S, Losurdo A, et al. Pain and the alpha-sleep anomaly: a mechanism of sleep disruption in facioscapulohumeral muscular dystrophy. *Pain Med* 2013 Apr 01;14(4):487-497. [doi: [10.1111/pme.12054](https://doi.org/10.1111/pme.12054)] [Medline: [23387524](https://pubmed.ncbi.nlm.nih.gov/23387524/)]
27. Runte M, Spiesshoefer J, Heidebreder A, Dreher M, Young P, Brix T, et al. Sleep-related breathing disorders in facioscapulohumeral dystrophy. *Sleep Breath* 2019 Apr 26;23(3):899-906. [doi: [10.1007/s11325-019-01843-1](https://doi.org/10.1007/s11325-019-01843-1)]
28. Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: a survey. *Pervasive Mobile Comput* 2018 Dec;51:1-26. [doi: [10.1016/j.pmcj.2018.09.003](https://doi.org/10.1016/j.pmcj.2018.09.003)]
29. Garcia-Ceja E, Galván-Tejada CE, Brena R. Multi-view stacking for activity recognition with sound and accelerometer data. *Inf Fusion* 2018 Mar;40:45-56. [doi: [10.1016/j.inffus.2017.06.004](https://doi.org/10.1016/j.inffus.2017.06.004)]
30. Soleymani M, Riegler M, Halvorsen P. Multimodal analysis of image search intent: intent recognition in image search from user behavior and visual content. In: *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*. 2017 Presented at: ICMR '17: International Conference on Multimedia Retrieval; Jun 6 - 9, 2017; Bucharest Romania. [doi: [10.1145/3078971.3078995](https://doi.org/10.1145/3078971.3078995)]
31. Andone I, Błaszczewicz K, Eibes M, Trendafilov B, Montag C, Markowetz A. How age and gender affect smartphone usage. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. 2016 Presented at: UbiComp '16: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 12 - 16, 2016; Heidelberg Germany. [doi: [10.1145/2968219.2971451](https://doi.org/10.1145/2968219.2971451)]

## Abbreviations

**CHDR:** Centre for Human Drug Research

**CHDR MORE:** Centre for Human Drug Research Monitoring Remotely

**FSHD:** facioscapulohumeral dystrophy

**LASSO:** least absolute shrinkage and selection operator

**MCC:** Matthews correlation coefficient

**TUG:** Timed Up-and-Go test

*Edited by A Mavragani; submitted 05.07.21; peer-reviewed by R Ciorap, C Wang, I Clay; comments to author 25.01.22; revised version received 06.04.22; accepted 11.04.22; published 13.09.22.*

*Please cite as:*

*Maleki G, Zhuparris A, Koopmans I, Doll RJ, Voet N, Cohen A, van Brummelen E, Groeneveld GJ, De Maeyer J*

*Objective Monitoring of Facioscapulohumeral Dystrophy During Clinical Trials Using a Smartphone App and Wearables: Observational Study*

*JMIR Form Res* 2022;6(9):e31775

URL: <https://formative.jmir.org/2022/9/e31775>

doi: [10.2196/31775](https://doi.org/10.2196/31775)

PMID: [36098990](https://pubmed.ncbi.nlm.nih.gov/36098990/)

©Ghobad Maleki, Ahnjili Zhuparris, Ingrid Koopmans, Robert J Doll, Noline Voet, Adam Cohen, Emilie van Brummelen, Geert Jan Groeneveld, Joris De Maeyer. Originally published in *JMIR Formative Research* (<https://formative.jmir.org/>), 13.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# A Versatile and Scalable Platform That Streamlines Data Collection for Patient-Centered Studies: Usability and Feasibility Study

Haley Huang<sup>1\*</sup>, BSc; Sofia Aschettino<sup>1</sup>, BSc; Nasim Lari<sup>1</sup>, PhD; Ting-Hsuan Lee<sup>2</sup>, MHS; Sarah Stothers Rosenberg<sup>2</sup>, RN, MSN, MPH; Xinyi Ng<sup>2</sup>, PhD; Stella Muthuri<sup>1</sup>, MPH, PhD; Anirudh Bakshi<sup>3</sup>, MSc; Korrin Bishop<sup>4</sup>, BSc; Hussein Ezzeldin<sup>2\*</sup>, PhD

<sup>1</sup>IBM Consulting, IBM, Bethesda, MD, United States

<sup>2</sup>Center for Biologics Evaluation and Research, United States Food and Drug Administration, Silver Spring, MD, United States

<sup>3</sup>Gevity Consulting Inc., Accenture, Ottawa, ON, Canada

<sup>4</sup>Korrin Bishop Writing & Editing, Kodak, TN, United States

\*these authors contributed equally

**Corresponding Author:**

Hussein Ezzeldin, PhD

Center for Biologics Evaluation and Research

United States Food and Drug Administration

10903 New Hampshire Avenue

Silver Spring, MD, 20993

United States

Phone: 1 240 402 8629

Email: [hussein.ezzeldin@fda.hhs.gov](mailto:hussein.ezzeldin@fda.hhs.gov)

## Abstract

**Background:** The Food and Drug Administration Center for Biologics Evaluation and Research (CBER) established the Biologics Effectiveness and Safety (BEST) Initiative with several objectives, including the expansion and enhancement of CBER's access to fit-for-purpose data sources, analytics, tools, and infrastructures to improve the understanding of patient experiences with conditions related to CBER-regulated products. Owing to existing challenges in data collection, especially for rare disease research, CBER recognized the need for a comprehensive platform where study coordinators can engage with study participants and design and deploy studies while patients or caregivers could enroll, consent, and securely participate as well.

**Objective:** This study aimed to increase awareness and describe the design, development, and novelty of the Survey of Health and Patient Experience (SHAPE) platform, its functionality and application, quality improvement efforts, open-source availability, and plans for enhancement.

**Methods:** SHAPE is hosted in a Google Cloud environment and comprises 3 parts: the administrator application, participant app, and application programming interface. The administrator can build a study comprising a set of questionnaires and self-report entries through the app. Once the study is deployed, the participant can access the app, consent to the study, and complete its components. To build SHAPE to be scalable and flexible, we leveraged the open-source software development kit, Ionic Framework. This enabled the building and deploying of apps across platforms, including iOS, Android, and progressive web applications, from a single codebase by using standardized web technologies. SHAPE has been integrated with a leading Health Level 7 (HL7®) Fast Healthcare Interoperability Resources (FHIR®) application programming interface platform, lupHealth, which allows participants to consent to 1-time data pull of their electronic health records. We used an agile-based process that engaged multiple stakeholders in SHAPE's design and development.

**Results:** SHAPE allows study coordinators to plan, develop, and deploy questionnaires to obtain important end points directly from patients or caregivers. Electronic health record integration enables access to patient health records, which can validate and enhance the accuracy of data-capture methods. The administrator can then download the study data into HL7® FHIR®-formatted JSON files. In this paper, we illustrate how study coordinators can use SHAPE to design patient-centered studies. We demonstrate its broad applicability through a hypothetical type 1 diabetes cohort study and an ongoing pilot study on metachromatic leukodystrophy to implement best practices for designing a regulatory-grade natural history study for rare diseases.

**Conclusions:** SHAPE is an intuitive and comprehensive data-collection tool for a variety of clinical studies. Further customization of this versatile and scalable platform allows for multiple use cases. SHAPE can capture patient perspectives and clinical data, thereby providing regulators, clinicians, researchers, and patient advocacy organizations with data to inform drug development and improve patient outcomes.

(JMIR Form Res 2022;6(9):e38579) doi:[10.2196/38579](https://doi.org/10.2196/38579)

## KEYWORDS

mobile app; patient experience data; data-collection app; mobile phone; usability; mHealth app; feasibility; user centered; eHealth; patient-generated data

## Introduction

### Barriers to the Study of Rare Diseases

Treatments for rare diseases comprise the fastest-growing subcategory of novel drugs and biologics approvals in the United States [1]. In 2020, 58% of all novel drugs approved by the US Food and Drug Administration (FDA) and 20% of all novel biological products were designated to treat rare diseases [2]. In the United States, a rare disease, as defined by the Orphan Drug Act of 1983, is a condition that affects <200,000 people in the United States [3]. However, despite significant progress in rare disease therapeutic development, *only a few hundred of the 7000 known rare diseases currently have an FDA-approved treatment* [4]. This leaves many patients with rare diseases with few or no treatment options.

Research and development for the treatment of rare diseases involve several challenges, including underdiagnosis and delayed diagnosis [5,6]. In addition, there is insufficient knowledge of the natural history of most rare diseases. Related to this is the heterogeneous pathogenesis of many rare diseases. Not all rare diseases present or progress in the same way from individual to individual, which complicates drug development [1].

The lack of information on the natural history of many rare diseases stems from the overall low prevalence of the diseases, making it difficult to recruit and retain sufficient population sizes for natural history studies and clinical trials. This includes gathering appropriate control groups against which to compare health outcomes in patients with rare diseases [7].

Furthermore, given the small population sizes available for these studies, other constraints arise. Patients are often widely dispersed geographically, which creates logistical barriers. Patients may have limited access to clinical sites where trials are conducted, which can increase the burden for those participating in the trials [8,9].

Owing to these challenges, clinical evidence on rare diseases is often limited because of a lack of sufficient and robust clinical data [10]. This negatively affects the quality of scientific evidence required to support clinical and regulatory decision-making.

An intuitive, comprehensive platform, such as the one described in this paper, presents an opportunity to confront these barriers and strengthen the clinical evidence base by making research on rare diseases more feasible. Specifically, it may be able to reduce the high burdens of travel, time, and overall fatigue faced

by patients and their caregivers when participating in rare disease studies at remote clinical sites. The Survey of Health and Patient Experience (SHAPE) participant app can easily collect patient experience data regardless of their location or time. Thus, the use of a mobile app could accelerate meeting the unmet needs of patients with rare diseases by helping researchers gain greater knowledge on disease conditions.

### Literature Review of Data-Collection Approaches for Rare Diseases

In recent years, innovative approaches have been developed offering substantial promise in promoting more efficient and effective ways of collecting data for epidemiological and clinical studies. These approaches can be applied to overcome methodological and logistical challenges inherent to rare disease research. A literature review was conducted to assess existing mobile apps and tools available for the collection of self-reported patient data used in clinical studies. A targeted search was conducted in PubMed by using the keywords *patient reported outcomes* and *patient reported outcome measures*. Additional details on this search strategy are included in [Multimedia Appendix 1](#).

The search identified numerous tools available for individual diseases or outcomes under study [11-20], as well as remappable resources (ie, migrating data from one system to another), such as REDCap (Research Electronic Data Capture; Vanderbilt University) and the FDA MyStudies app [21,22]. Most single-study apps are limited to the disease or condition under investigation, such as Parkinson disease and asthma [11,12]. The FDA MyStudies app collects information that is unavailable in their electronic health records (EHRs) directly from patients and combines it with the data from their EHRs [22]. REDCap is a web-based application used to create surveys and research forms to capture data for clinical research [21]. These tools were assessed to understand the current state of the field regarding the existing functionality, applications, strengths, and gaps.

### Purpose of the SHAPE Platform

The FDA Center for Biologics Evaluation and Research (CBER) Biologics Effectiveness and Safety (BEST) Initiative recognizes the need for innovative approaches that facilitate efficient and effective ways of collecting fit-for-purpose data for patient-centered studies. To respond to this need, the Biologics Effectiveness and Safety Initiative sought to design and develop the SHAPE platform.

SHAPE was developed to provide an open-source platform that study coordinators and research teams can use to customize

studies and scale data-collection efforts. It was designed with interoperability in mind to allow seamless health data exchange and integration with EHR systems while meeting the FDA's regulatory requirements regarding data security and traceability. SHAPE also facilitates the direct data capture of patient health experiences with a range of health conditions, symptoms, and health care services, as well as the collection of end points that may be relevant to regulators, clinicians, researchers, and patient advocacy organizations. These capabilities allow the collection of longitudinal data outside traditional onsite methods while reducing the burden that patients and their caregivers may face when participating in conventional studies at remote clinical sites, such as travel, costs, time, and overall fatigue. SHAPE is intuitive and highly customizable and offers a wide range of functionalities (eg, questionnaires, self-reports, and EHR linkage), providing a cost-effective solution for organizations to host multiple studies with little additional cost while maintaining control of the platform. SHAPE's codebase is shared on the FDA public GitHub and is available to any organization interested in leveraging the codebase to set up their own instance of SHAPE [23].

The objective of this study was to describe the design, development, and novelty of SHAPE; its functionality and

applications; quality improvement efforts; and plans for its enhancement.

## Methods

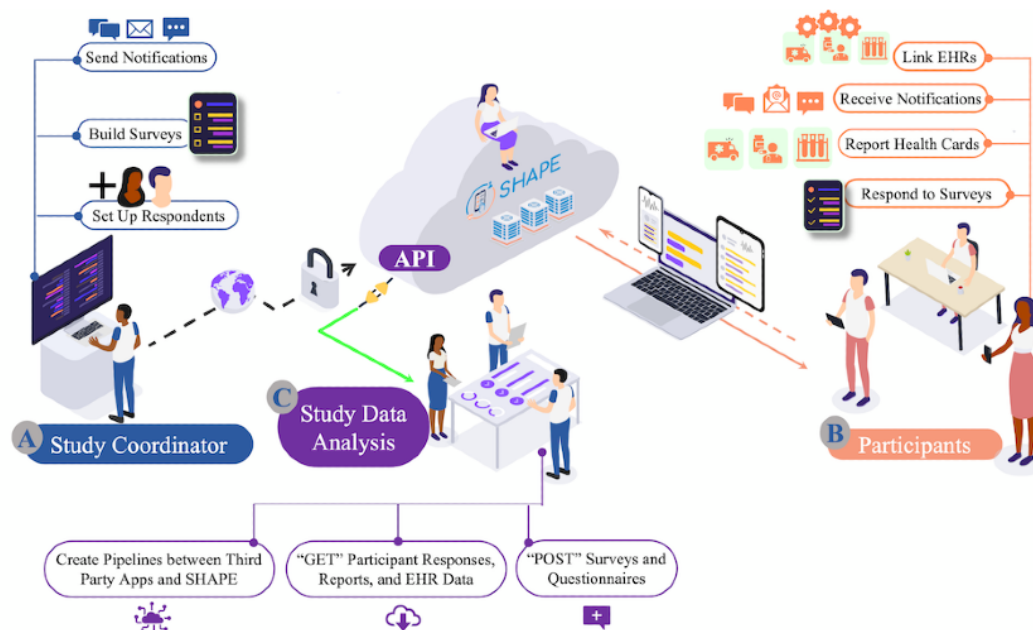
### Overview

SHAPE is a user-friendly platform that teams can leverage to customize studies and scale data-collection efforts to advance health research. It comprises the following three parts:

1. *Administrator application* that study coordinators and administrators can use to develop and deploy study questionnaires and share important information and resources with participants, shown as A in Figure 1
2. *Participant app* that allows research participants to explicitly consent to and participate in patient-centered studies through their mobile device, shown as B in Figure 1
3. *Application programming interface (API)* that study coordinators can use to push and pull data to SHAPE, shown as C in Figure 1

Figure 1 shows the overall SHAPE architecture, highlighting major features and displaying the various SHAPE users and how they interact with the platform.

**Figure 1.** SHAPE overall architecture. API: application programming interface; EHR: electronic health record; SHAPE: Survey of Health and Patient Experience.



### Definitions

Throughout this paper, the following terms refer to the corresponding elements of the SHAPE platform:

1. *SHAPE* refers to the platform as a whole.
2. *SHAPE administrator application* refers to the application component used by study coordinators and other administrators.
3. *SHAPE participant app* refers to the app component used by patients and caregivers participating in a SHAPE-based study.
4. *SHAPE API* refers to the API study coordinators can use.

In addition, we use the following terms to further describe the platform and explain the different use scenarios:

1. *Participant* refers to the individual completing a questionnaire in the SHAPE participant app. This can be either a patient or a caregiver who is responding on the patient's behalf. For instances where it is important to make a distinction, we use *patient* or *caregiver* to identify the respondent specifically.
2. *Study* refers to the questionnaires and health events developed with SHAPE to research a specific question about patient experiences. Although SHAPE uses the term *survey* within the platform, we have chosen to use the term



*study* throughout this paper to clarify that research teams can use SHAPE not just for surveys but also for developing comprehensive studies, including longitudinal studies that may use multiple questionnaires over time.

3. *Questionnaire* refers to the individual sets of questions that study administrators can create within SHAPE for participants to answer.

## Design and Development

### *Agile Application Development Process*

The SHAPE team comprised members from the FDA CBER, IBM, and National Organization for Rare Disorders (NORD) and patient advocates and caregivers connected to NORD. This team used an agile-based software development process, using a combination of principles, methodologies, and delivery expertise. The team worked closely with other subject matter experts to gain a deep understanding of the outcome vision for SHAPE, as well as the requirements and organizational processes needed.

Using an agile-based approach for developing SHAPE had many benefits. An emphasis on stakeholder engagement allowed the team to work with regulators, caregivers, study coordinators, and clinicians to understand and address their current challenges. This included greater insights into the limitations of current methods for recruiting and interacting with patients. The team was able to use stakeholder knowledge to develop features and designs to meet their needs.

The SHAPE team designed, developed, and implemented functionality in 2-week cycles, completing >60 sprints to date.

### *SHAPE Architecture*

To build SHAPE as a scalable and flexible platform, the team used an open-source software development kit—Ionic Framework [24]—to build the various SHAPE components and leverage its easy deployment functionality. This allowed the team to deploy SHAPE across multiple operating systems, including iOS, Android, and progressive web applications from a single codebase, using standardized web technologies.

SHAPE is hosted in a Google Cloud environment. This cloud environment provides SHAPE with a secure infrastructure that is scalable while requiring minimal associated costs. These features allow support for large user base populations.

The SHAPE participant app is available in both the Apple App Store and Google Play Store for user downloads. Most updates to the SHAPE participant app are automatically pushed to the users' mobile devices without the need for a user-initiated upgrade through the app stores. If the development team implements a major feature that is considered a binary update, app users must update the SHAPE participant app via their operating system's app store.

Currently, studies in SHAPE are by invitation only. Study administrations provide participants with an ID to be used during registration through the SHAPE participant app. To enroll participants in any study, the administrator has two options:

1. Batch input participants by uploading a CSV file containing unique, non-personally identifiable codes

2. Manually enter this information by using a participant import wizard within the SHAPE administrator application

The SHAPE administrator application enables study administrators to build a study, upload participant study cohorts, and communicate with the cohorts through 1 interface. The study built on the SHAPE administrator app is then deployed to participants through the SHAPE participant app.

Each study comprises a set of questionnaires and self-reports. Self-reports are a set number of predetermined questions—curated by study administrators—that are always accessible to participants for ad hoc data submission. These can be health event questions on the occurrence of a medical diagnosis (eg, infection), symptom (eg, fatigue), or medical visit (eg, emergency room or primary care visit). Participants can report these at any point during the time frame of the study by using the SHAPE participant app.

SHAPE is secure and restricted to only administrators and participants of official studies leveraging SHAPE for data-collection efforts. As a result, the apps require authentic log-in credentials to gain access. To ensure a secure connection, the team has developed the SHAPE API, which allows study administrators with approved credentials to automatically push study-specific data to the SHAPE administrator application and pull collected participant data by developing their own code that hits the API target end point. Partner platforms can set up a seamless integration connection if the user is authenticated to connect to the end points.

SHAPE's flexible and scalable solution architecture allows for administrator multitenancy. This restricts access for accounts based on their organization, enabling users to only view and export their organization's studies and participant data. All organizations must work with the SHAPE team to request an administrator account, which will create their own organization-specific sandbox. A benefit of the multitenancy feature is that it allows for multiple groups within an organization to leverage SHAPE without requiring the development and maintenance of their own instance of SHAPE. The cost and time savings are substantial, as the groups will not need to adopt and build the app for each new study.

## Functionality

### *Fast Healthcare Interoperability Resources and EHR Integration*

SHAPE supports the use of the Health Level 7 (HL7®) Fast Healthcare Interoperability Resources (FHIR®) standard in its data export functionality. FHIR® is a leading innovative standard in the health care field, as it facilitates easier exchange and use of data for research by providing a common way of defining and representing exchangeable content (ie, resources) and a common set of metadata. SHAPE's data exports are JSON files. Administrators can download the JSON data export in two formats: (1) typical JSON following no data standard or (2) JSON with data mapped to HL7® FHIR® data standards.

SHAPE has been integrated with a third-party application, lupHealth, to provide participants with the ability to approve a 1-time data pull of their EHRs into the study. This is an

innovative feature for SHAPE as participants' EHRs are not readily accessible. SHAPE connects to IupHealth's platform and directs participants to sign into their medical providers' EHR portals. Once logged in, the participant is provided with the ability to consent to download their health records in FHIR®-formatted files and initiate a 1-time data pull of that information into SHAPE. This ultimately gives the study administrator access to the information.

The benefits of this feature are high for both the participants and study administrators. Researchers can use EHR data to expand data capture and improve the identification of medical conditions, symptoms, current medications, past treatment history, and so on. The addition of EHR data can also minimize missing data and potential recall bias from self-reported medical history data.

From the perspective of the study participants, they have increased transparency into their health records. This occurs through the provision of a 1-time readable version of EHR data that their provider has in their database at the time of import. As a result, they can view their health records in conjunction with their own reported data to facilitate informed conversations with their clinician or clinicians on medical plans. This feature supports the increased importance of exchanging information between health care providers, their patients, and the regulators tasked with ensuring the safety and effectiveness of medications, vaccines, and therapies. The 2 data sources, patient-submitted data and their EHR data, complement each other for a more comprehensive picture of a patient's medical needs and care.

### ***Study Design and Development***

This section and the sections that follow include screenshots from a live instance of SHAPE. The screenshots include hypothetical questions developed to demonstrate how administrators can use SHAPE to gather data from patients. All data displayed in this paper are simulated and are not personally identifiable.

For a study to meet institutional review board requirements and properly protect its participants, study administrators must obtain informed consent from each participant. SHAPE provides for efficient collection of these requirements. The SHAPE participant app will only let participants join a study after they have reviewed the terms listed for the study and provided their informed consent within the app. Once the participant has viewed and agreed to the terms, SHAPE delivers a copy of the

informed consent to their provided email address and then displays the study to them. SHAPE also tracks the consent in the back-end database for the study's own audit purposes.

Currently, SHAPE is a secure platform that only invited study participants can access. The administrator shares a unique code with each participant outside the app. The participant then uses the provided code to register with the SHAPE participant app. Once the study is developed and deployed, the participant can access it via the SHAPE participant app, consent to the study, and complete any questionnaires or self-reported events related to their disease throughout the study's time frame. During the study, the administrator can communicate with participants through in-app messages, emails, and SMS text messaging. If enabled by device owners, the SHAPE participant app can also connect to the device's native notification center and deliver push notifications to users.

SHAPE supports many of the main modality questions leveraged for studies, such as single-line text, radio button, date and time, checkbox, dropdown, slider, text area, range (limits the minimum and maximum integers a respondent can enter), and information only (there is no response field for the participants; instead, administrators can display images, tables, and text for information-sharing purposes).

When developing questions, study administrators can use complex skip logic rules to ensure that their questionnaire captures all relevant data while only displaying questions applicable to the individual participant. SHAPE supports skip logic rules that can be developed according to the participant's gender, age, or response to the posed question.

The SHAPE administrator application can also provide administrators with a preview of questionnaires in development. This functionality enables administrators to test the questionnaire's skip logic while they are building it to ensure that it flows correctly for the participants.

Study administrators can target SHAPE's self-report functionality to capture patient data on potential health events, results of a clinical visit, or requests to withdraw from the study. The participant can open and submit an entry at their own convenience, in addition to the standard questionnaires deployed within a study. Once a self-report diary entry is submitted, SHAPE holds a receipt of the data collected and displays the information provided to the participant in their diary history view (Figure 2).

**Figure 2.** Example self-report in the Survey of Health and Patient Experience app for a health event.

9:44

Self-Report

< Back

**Health Event 2022-02-18T12:11:51-05:00**

Event Recorded Date: 2022-02-18T12:11:51-05:00

Description: emergency

End Date: 01/20/2022

Event Treatment: Medication

Health Event: Emergency Department Visit

Health Event Specification:

Ongoing Status: No

Onset Date: 01/18/2022

Outcome: Recovered

Outcome:

Post Event Treatment: None

Date of Birth: 08/13/1996

Name of Participant: Jack

Surveys Profile Messages Self-Report

## Results

### Use Scenario

To illustrate how study coordinators can use SHAPE to plan and design patient experience studies, we developed a hypothetical type 1 diabetes cohort study. By using SHAPE, a study coordinator can develop a cohort population based on patients who have signed up for the study. SHAPE can support a study that focuses on obtaining treatment and clinical end points across a series of patient groups while also obtaining longitudinal data through the deployment of a series of questionnaires (Figure 3).

In our hypothetical study, we were interested in understanding treatment patterns and outcomes of patients with diabetes. In

this scenario, the study coordinator can develop questionnaires that gather baseline information and health data, followed by a series that captures key treatment end points at the 3-, 6-, and 12-month time points of the study.

For example, the baseline survey may collect key health and treatment information, such as the names of participants' primary physicians or specialists, when they were diagnosed with type 1 diabetes, their hemoglobin A<sub>1c</sub> levels, and other comorbidities. Figure 4 illustrates a selection of these supported baseline SHAPE question types.

Study coordinators can then create a series of health event diary forms that participants can submit at their leisure. These can capture any health event updates or other key ongoing clinical end points of interest (Figure 5).

Figure 3. Questionnaire view in the Survey of Health and Patient Experience participant app.

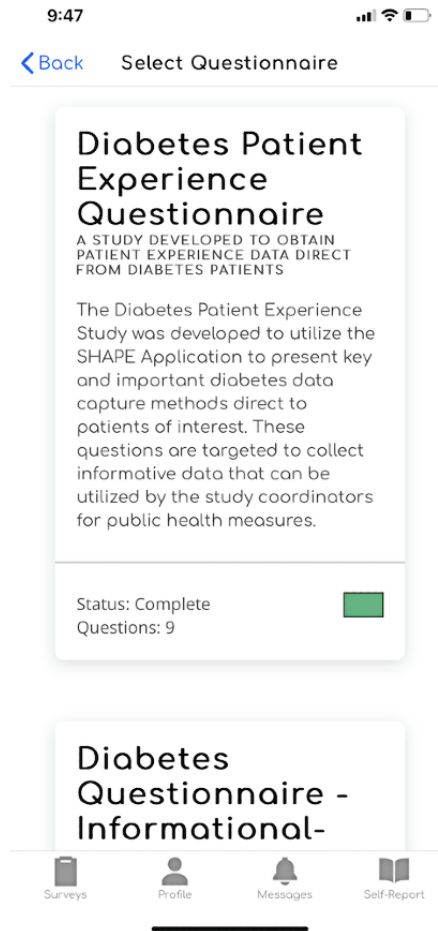
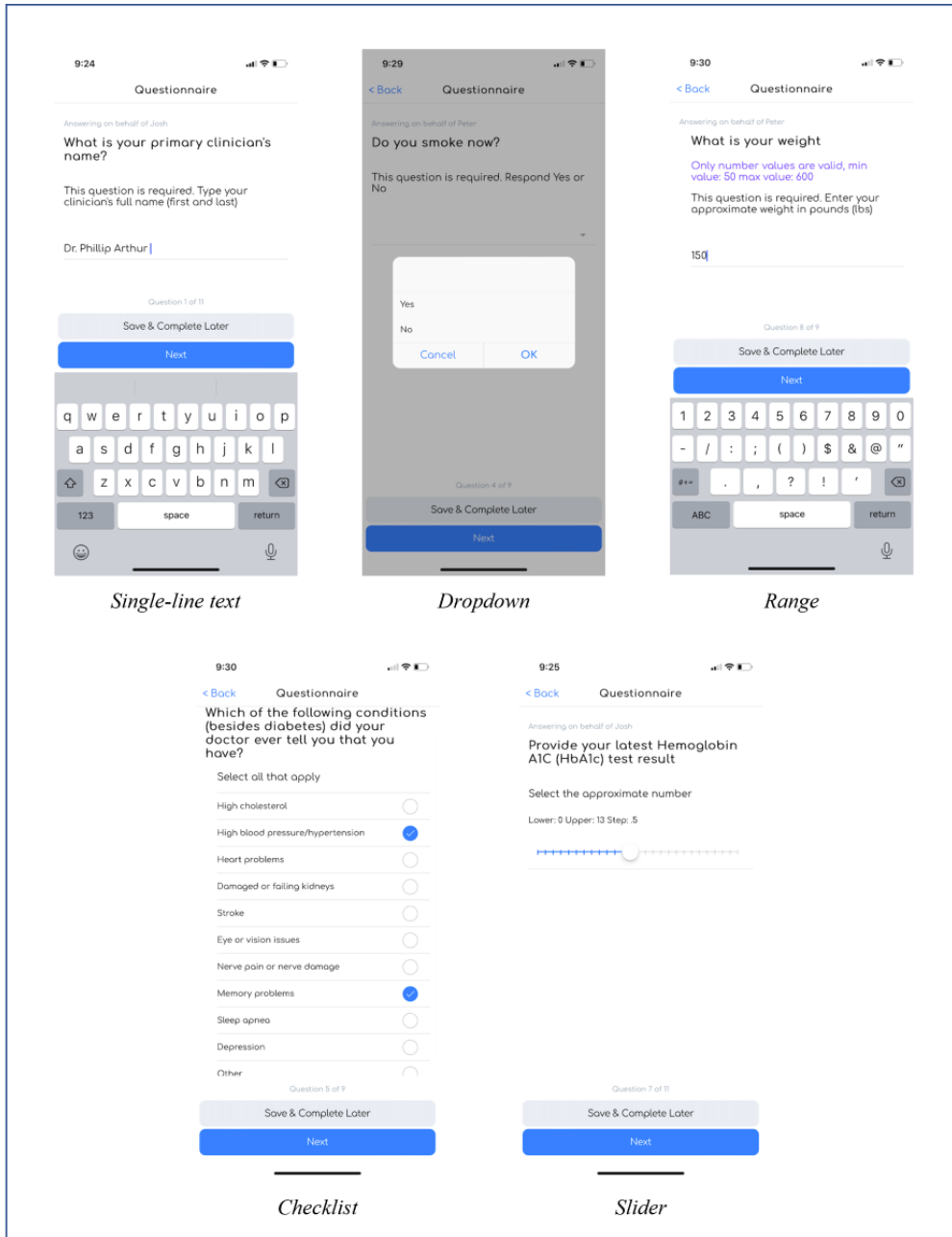


Figure 4. Questionnaire view in the Survey of Health and Patient Experience app question types.



**Figure 5.** Health event self-report screenshots.

When a questionnaire is ready to be deployed, the study coordinator has the option to send a push notification (Figure 6) to all mobile app users who have installed the SHAPE participant app and signed up for the study. The coordinator can also develop a message to send via email, SMS text messaging, or the in-app messaging system to inform the study participants of the task.

Study coordinators can deliver a message inviting participants interested in linking their EHRs via the SHAPE participant app to share additional data with the study coordinators (Figure 7). They can then use these data to confirm a patient's diabetes diagnosis and note any additional information of interest, such as other conditions, medications, or procedures. Access to EHR data allows study coordinators to validate the information from patients' self-reported health events, addressing potential issues with patient recall by not needing to rely solely on self-reported data.

Following baseline questionnaire data collection, the study coordinator is ready to deploy the treatment-specific patient experience questionnaires for longitudinal updates from the 2

treatment populations across the full length of the study. The treatment-specific questionnaires (Figures 8 and 9) can be designed to capture key treatment patient experience end points around use, accessibility, burden, and key vital measurements (eg, hemoglobin A<sub>1c</sub> level and weight).

In addition to developing questions for important end point data collection, study coordinators can use the *information only* question type (Figure 10) to share important information or resources with the participants.

SHAPE is an intuitive and comprehensive data-collection tool for a variety of clinical studies. It allows study coordinators to plan, curate, develop, and deploy their questionnaires to obtain important end points directly from patients or through caregivers. The EHR integration functionality enables greater access to the patient's health records while protecting patient privacy. This allows patients to see their health records at the point of import and provides research coordinators with the use of the data for validation and additional data-capture methods. Overall, this provides a more comprehensive picture of patient experiences with the condition or therapeutic being studied.

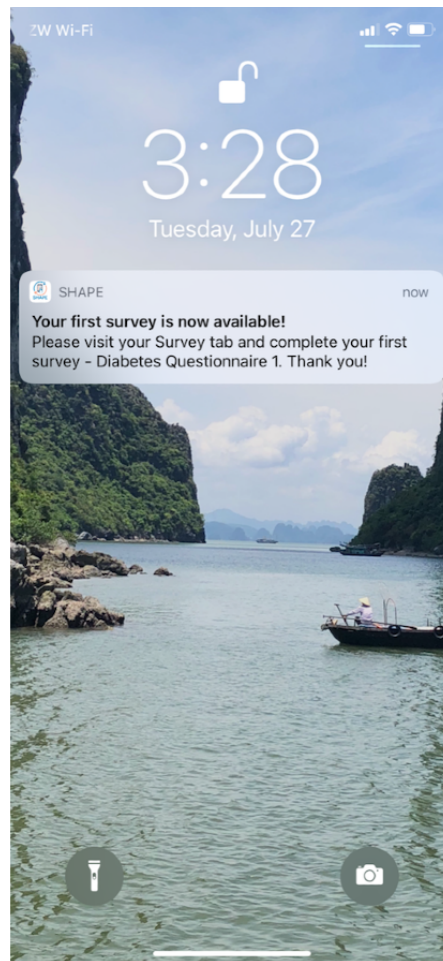
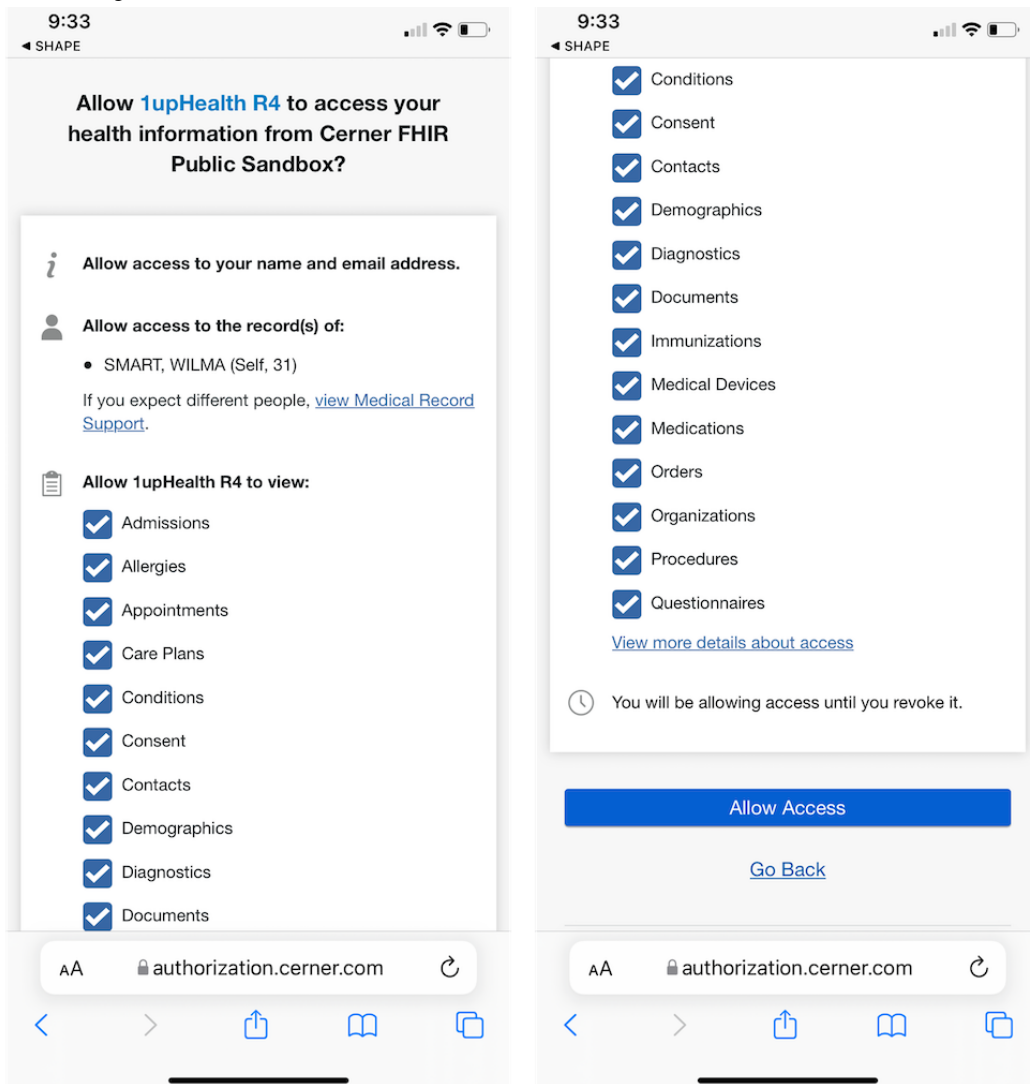
**Figure 6.** Example push notification on a mobile device.

Figure 7. Process for allowing links to the electronic health record.





**Figure 8.** Treatment-specific question examples that capture the use, accessibility, and burden.

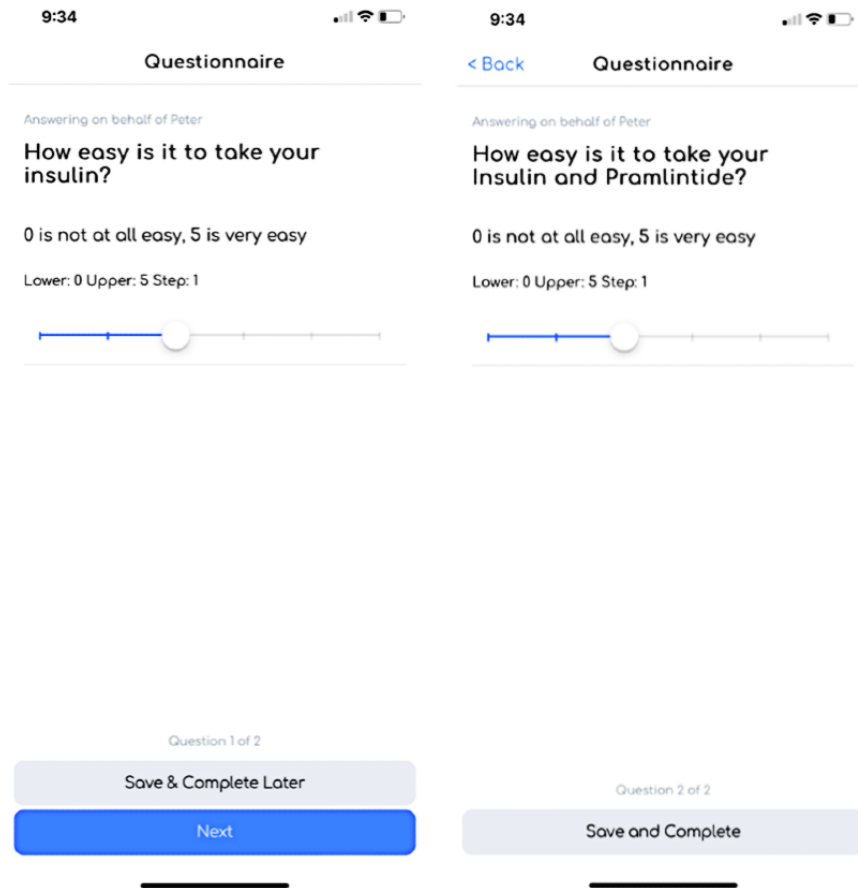
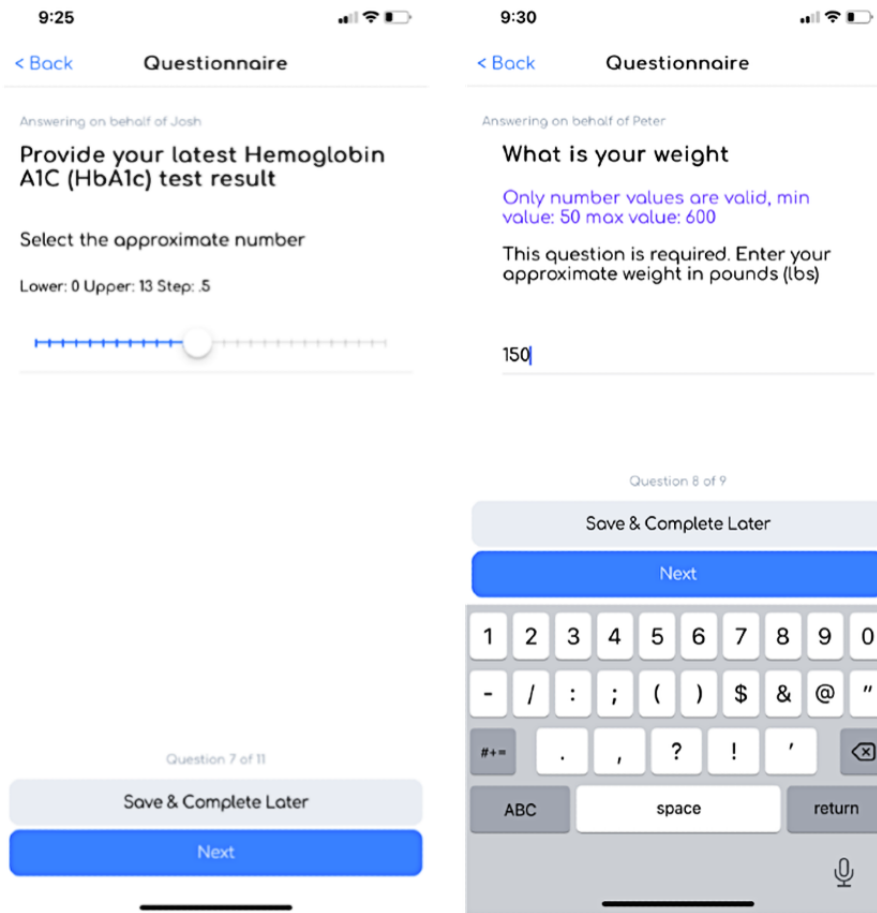
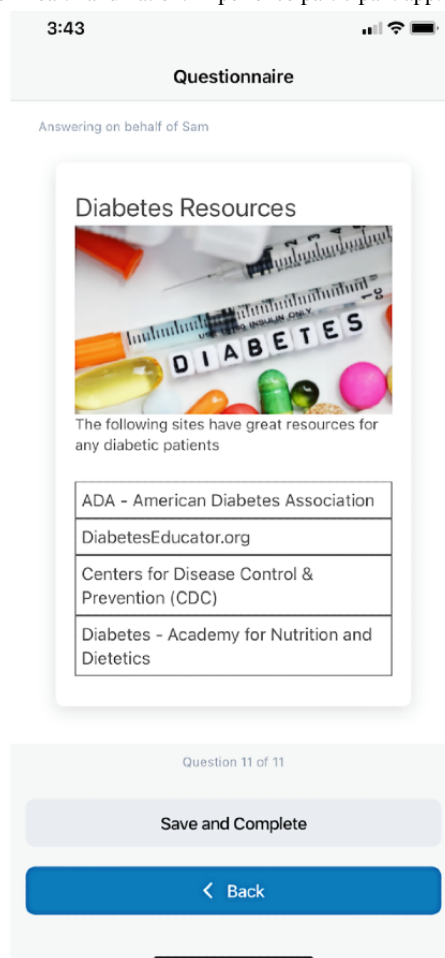


Figure 9. Treatment-specific question examples that capture key vital measurements.



**Figure 10.** Information only page in the Survey of Health and Patient Experience participant app.

### In-Progress Use Case

To address the unmet needs in rare disease therapeutics, the FDA CBER, in collaboration with NORD, launched a pilot Natural History of Metachromatic Leukodystrophy study to implement best practices and processes for designing a regulatory-grade natural history study for rare diseases [25].

This study required a user-friendly app for the deployment of patient experience questionnaires for collecting prespecified study end points. The study is currently in progress in patients with metachromatic leukodystrophy using the SHAPE participant app to provide patient experience data to NORD study coordinators [26,27].

## Discussion

### Principal Findings

As part of the FDA's mission to facilitate data collection and understanding of patient experiences, SHAPE offers easy-to-use and intuitive SHAPE administrator application and participant app that are appropriate for a variety of studies. In its pilot deployment, SHAPE was used to collect data relevant to regulatory decision-making and patient health outcomes. It aimed to address common challenges of attrition and missing data in natural history studies. The platform allows participants to enter important end point data and other episodic health events at any point during the study from any place with internet

access. They can use a computer at their home or submit data on the go with their mobile devices through the SHAPE participant app.

### Strengths

SHAPE has several strengths. First, to use SHAPE, study coordinators are not required to have technical expertise in software development or programming to build, deploy, or collect information from the platform. A study coordinator can log into the SHAPE administrator application and within its 1 interface be able to create studies, incorporate the study's informed consent, upload participant IDs, curate questionnaires, send communications, view data, export data, and actively open and close participant studies. This offers a low barrier of entry for research teams. SHAPE was designed with the user experience in mind, making it an accessible, flexible, and scalable data-collection platform for both research participants and study coordinators.

In addition, SHAPE allows for easy scalability through its multitenancy ability, which allows for multiple organizations to use SHAPE with their own secure log-ins. This ensures that their studies are secure and restricts access for accounts based on their organization, enabling users to only view and export their organization's studies and participant data. This feature permits multiple groups within an organization to leverage SHAPE without requiring the development of their own instance of the platform. This leads to substantial cost and time savings.

Finally, SHAPE can potentially reduce the burdens patients and their caregivers may face while participating in conventional regulatory-grade studies at remote clinical sites, such as travel, costs, time, and overall fatigue. It facilitates an easier exchange and use of data for patient-centered research. For example, the EHR integration function allows research participants to explicitly consent to share their EHR data in a deidentified manner, making it a more convenient procedure for data capture while improving the accuracy of collected data and minimizing missing data.

### Limitations

A limitation of SHAPE is that it requires individuals to have access to the internet or smartphones or tablets. It may not be suitable for individuals who are uncomfortable or unfamiliar with using technology. Currently, SHAPE is only available in the English language and lacks accessibility features for individuals who have impaired vision or dexterity (eg, people with Parkinson disease). Future enhancements of the platform could address some of these limitations.

### Comparison With Prior Work

There are several open-source mobile data-collection platforms that support data management for clinical research [11-20]. Many of these serve as important functions and played a significant role in the design and development of SHAPE in terms of its capabilities, functionality, and application. However, few platforms have adequately addressed informed consent processes, such as data security and traceability, that research studies intended for regulatory use require [28]. An exception is the FDA MyStudies app, which was created to conform to stringent federal data privacy and security standards.

SHAPE builds on the prior capabilities of the MyStudies app and includes additional important features and applications, such as customizable questionnaires, self-report, EHR integration capabilities, communication channels, and compatibility with multiple operating systems. These enhancements provide greater ease of use for study coordinators and participants. SHAPE is also unique in its ability to support the use of FHIR® and seamlessly integrate with EHR systems. These capabilities make SHAPE a scalable and flexible data-collection platform that research teams can use to deploy multiple studies without requiring additional configuration by software developers.

Finally, the initial version of the SHAPE platform had good functionalities to support small study groups such as those connected to rare disease research but lacked the automation of studies to support larger cohorts. As SHAPE continues to develop, it will gain broader applicability for research for prevalent diseases. Further automation will be able to support larger study cohorts while keeping costs for study administrators low and study tools easily accessible for study participants. Although rare disease research served as a vital gateway to the development of SHAPE, the platform's utility can easily expand from rare disease studies to regulatory-grade research for a broad range of diseases.

### Quality Improvement and Future Plans for Enhancement

The SHAPE team has developed guides and training documentation on how to use the SHAPE participant app, SHAPE administrator application, and SHAPE API. The team adds updates to these resources regularly to clarify elements based on user feedback and to highlight upgrades to the tool.

As part of the app's ongoing quality improvement, the SHAPE team will continue to engage all relevant business and user stakeholders to discuss and prioritize updates. On the basis of the feedback and initial use of the platform, the team has already determined several plans for enhancement.

Future feature development for SHAPE is targeted at increasing the customizability and usability of the tool for the administrators to reduce the study administration burden. Updates will also continually advance SHAPE's ability to integrate into the native functionalities of a smartphone mobile device, including the camera, calendar, and file records. In addition, future enhancements will further strengthen the platform's communication functionality and ability to schedule reminders for questionnaires and self-report submissions over set periods, such as every 3 months. This will support longitudinal data capture. There are also plans to provide study administrators with the ability to add respondents to a secondary questionnaire based on a provided response. In addition, the team plans to expand participant registration beyond invites only, allowing app users to find and register for relevant public studies on their own.

Finally, the SHAPE team would like to improve the EHR feature to populate the participant's consented data collection into a display that is easy to read and that participants can export and print for record-keeping purposes. Future iterations of SHAPE will also position it as an information-sharing channel with a repository of resources that can support clinical trials, study participants, and so on.

### Conclusions

SHAPE is a flexible and scalable app that allows for efficient deployment of regulatory-grade studies, supports multitenancy features, streamlines access to EHR data, eliminates the need for ongoing software developer support, and so on.

This intuitive and comprehensive data-collection platform can be used to support traditional data-capture methods from clinical trials, pragmatic trials, observational studies, and disease registries. To increase the ease of study participation and collect patient experience information, the FDA CBER used SHAPE for a pilot natural history study for a rare disease.

Further customization and automation of this versatile and scalable app will allow for multiple use cases, including larger cohort studies for common diseases. Study administrators can use SHAPE to capture patient perspectives and important clinical data. This provides a more holistic view of each patient, facilitating the goal of achieving better patient outcomes through drug development and delivery of care.

## Acknowledgments

Development of the manuscript benefited from significant engagement with the Food and Drug Administration Center for Biologics Evaluation and Research team members and their partners. The authors thank them for their contributions and feedback. Additional feedback on the manuscript was provided by IBM Consulting (Lance Jones and Michael Griffith); Gevity Consulting Inc, a part of Accenture (Patrick Saunders-Hastings); and Epi Excellence LLC (Pradeep Rajan). SSR conducted this study while affiliated with the Center for Biologics Evaluation and Research (CBER), the author is now affiliated with the FDA Center for Drug Evaluation and Research (CDER).

## Conflicts of Interest

HH, SA, NL, and SM are employed by IBM. AB is employed by Gevity Consulting Inc, a part of Accenture.

## Multimedia Appendix 1

Literature review search strategies.

[[DOCX File, 42 KB - formative\\_v6i9e38579\\_app1.docx](#)]

## References

1. Kesselheim AS, Gagne JJ. Strategies for postmarketing surveillance of drugs for rare diseases. *Clin Pharmacol Ther* 2014 Mar;95(3):265-268. [doi: [10.1038/clpt.2013.218](https://doi.org/10.1038/clpt.2013.218)] [Medline: [24193169](https://pubmed.ncbi.nlm.nih.gov/24193169/)]
2. Rare Disease Day 2021: FDA Shows Sustained Support of Rare Disease Product Development During the Public Health Emergency. U.S. Food & Drug Administration. 2021. URL: <https://www.fda.gov/news-events/fda-voices/rare-disease-day-2021-fda-shows-sustained-support-rare-disease-product-development-during-public> [accessed 2022-03-17]
3. Orphan Drug Act - Relevant Excerpts. U.S. Food & Drug Administration. 2013. URL: <https://www.fda.gov/industry/designating-orphan-product-drugs-and-biological-products/orphan-drug-act-relevant-excerpts> [accessed 2022-03-17]
4. Barriers to Rare Disease Diagnosis, Care, and Treatment in the US: A 30-Year Comparative Analysis. National Organization for Rare Diseases. 2020 Nov 19. URL: [https://rarediseases.org/wp-content/uploads/2020/11/NRD-2088-Barriers-30-Yr-Survey-Report\\_FNL-2.pdf](https://rarediseases.org/wp-content/uploads/2020/11/NRD-2088-Barriers-30-Yr-Survey-Report_FNL-2.pdf) [accessed 2022-03-17]
5. Molster C, Urwin D, Di Pietro L, Fookes M, Petrie D, van der Laan S, et al. Survey of healthcare experiences of Australian adults living with rare diseases. *Orphanet J Rare Dis* 2016 Mar 24;11:30 [FREE Full text] [doi: [10.1186/s13023-016-0409-z](https://doi.org/10.1186/s13023-016-0409-z)] [Medline: [27012247](https://pubmed.ncbi.nlm.nih.gov/27012247/)]
6. Senior TP, Knight AW. Rare diseases need a generic approach. *BMJ* 2006 Sep 16;333(7568):602 [FREE Full text] [doi: [10.1136/bmj.333.7568.602-a](https://doi.org/10.1136/bmj.333.7568.602-a)] [Medline: [16974023](https://pubmed.ncbi.nlm.nih.gov/16974023/)]
7. Nestler-Parr S, Korchagina D, Toumi M, Pashos CL, Blanchette C, Molsen E, et al. Challenges in research and health technology assessment of rare disease technologies: report of the ISPOR rare disease special interest group. *Value Health* 2018 May;21(5):493-500 [FREE Full text] [doi: [10.1016/j.jval.2018.03.004](https://doi.org/10.1016/j.jval.2018.03.004)] [Medline: [29753344](https://pubmed.ncbi.nlm.nih.gov/29753344/)]
8. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, Rare Diseases Clinical Research Network. Clinical research for rare disease: opportunities, challenges, and solutions. *Mol Genet Metab* 2009 Jan;96(1):20-26 [FREE Full text] [doi: [10.1016/j.ymgme.2008.10.003](https://doi.org/10.1016/j.ymgme.2008.10.003)] [Medline: [19013090](https://pubmed.ncbi.nlm.nih.gov/19013090/)]
9. Augustine EF, Adams HR, Mink JW. Clinical trials in rare disease: challenges and opportunities. *J Child Neurol* 2013 Sep;28(9):1142-1150 [FREE Full text] [doi: [10.1177/0883073813495959](https://doi.org/10.1177/0883073813495959)] [Medline: [24014509](https://pubmed.ncbi.nlm.nih.gov/24014509/)]
10. Gagne JJ, Thompson L, O'Keefe K, Kesselheim AS. Innovative research methods for studying treatments for rare diseases: methodological review. *BMJ* 2014 Nov 24;349:g6802 [FREE Full text] [doi: [10.1136/bmj.g6802](https://doi.org/10.1136/bmj.g6802)] [Medline: [25422272](https://pubmed.ncbi.nlm.nih.gov/25422272/)]
11. 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 [FREE Full text] [doi: [10.1038/sdata.2016.11](https://doi.org/10.1038/sdata.2016.11)] [Medline: [26938265](https://pubmed.ncbi.nlm.nih.gov/26938265/)]
12. Chan YF, Bot BM, Zweig M, Tignor N, Ma W, Suver C, et al. The asthma mobile health study, smartphone data collected using ResearchKit. *Sci Data* 2018 May 22;5:180096 [FREE Full text] [doi: [10.1038/sdata.2018.96](https://doi.org/10.1038/sdata.2018.96)] [Medline: [29786695](https://pubmed.ncbi.nlm.nih.gov/29786695/)]
13. Chan YF, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. *Nat Biotechnol* 2017 Apr;35(4):354-362 [FREE Full text] [doi: [10.1038/nbt.3826](https://doi.org/10.1038/nbt.3826)] [Medline: [28288104](https://pubmed.ncbi.nlm.nih.gov/28288104/)]
14. Crouthamel M, Quattrocchi E, Watts S, Wang S, Berry P, Garcia-Gancedo L, et al. Using a ResearchKit smartphone app to collect rheumatoid arthritis symptoms from real-world participants: feasibility study. *JMIR Mhealth Uhealth* 2018 Sep 13;6(9):e177 [FREE Full text] [doi: [10.2196/mhealth.9656](https://doi.org/10.2196/mhealth.9656)] [Medline: [30213779](https://pubmed.ncbi.nlm.nih.gov/30213779/)]
15. Egger HL, Dawson G, Hashemi J, Carpenter KL, Espinosa S, Campbell K, et al. Automatic emotion and attention analysis of young children at home: a ResearchKit autism feasibility study. *NPJ Digit Med* 2018 Jun 1;1:20 [FREE Full text] [doi: [10.1038/s41746-018-0024-6](https://doi.org/10.1038/s41746-018-0024-6)] [Medline: [31304303](https://pubmed.ncbi.nlm.nih.gov/31304303/)]
16. Radin JM, Steinhubl SR, Su AI, Bhargava H, Greenberg B, Bot BM, et al. The Healthy Pregnancy Research Program: transforming pregnancy research through a ResearchKit app. *NPJ Digit Med* 2018 Sep 5;1:45 [FREE Full text] [doi: [10.1038/s41746-018-0052-2](https://doi.org/10.1038/s41746-018-0052-2)] [Medline: [31304325](https://pubmed.ncbi.nlm.nih.gov/31304325/)]

17. Webster DE, Suver C, Doerr M, Mounts E, Domenico L, Petrie T, et al. The Mole Mapper Study, mobile phone skin imaging and melanoma risk data collected using ResearchKit. *Sci Data* 2017 Feb 14;4:170005 [FREE Full text] [doi: [10.1038/sdata.2017.5](https://doi.org/10.1038/sdata.2017.5)] [Medline: [28195576](https://pubmed.ncbi.nlm.nih.gov/28195576/)]
18. Yamaguchi S, Waki K, Nannya Y, Nangaku M, Kadowaki T, Ohe K. Usage patterns of GlucoNote, a self-management smartphone app, based on ResearchKit for patients with type 2 diabetes and prediabetes. *JMIR Mhealth Uhealth* 2019 Apr 24;7(4):e13204 [FREE Full text] [doi: [10.2196/13204](https://doi.org/10.2196/13204)] [Medline: [31017586](https://pubmed.ncbi.nlm.nih.gov/31017586/)]
19. Yoshimura Y, Ishijima M, Ishibashi M, Liu L, Arikawa-Hirasawa E, Machida S, et al. A nationwide observational study of locomotive syndrome in Japan using the ResearchKit: the Locomonitor study. *J Orthop Sci* 2019 Nov;24(6):1094-1104. [doi: [10.1016/j.jos.2019.08.009](https://doi.org/10.1016/j.jos.2019.08.009)] [Medline: [31492535](https://pubmed.ncbi.nlm.nih.gov/31492535/)]
20. Zens M, Woias P, Suedkamp NP, Niemeyer P. "Back on track": a mobile app observational study using Apple's ResearchKit framework. *JMIR Mhealth Uhealth* 2017 Feb 28;5(2):e23 [FREE Full text] [doi: [10.2196/mhealth.6259](https://doi.org/10.2196/mhealth.6259)] [Medline: [28246069](https://pubmed.ncbi.nlm.nih.gov/28246069/)]
21. Patridge EF, Bardyn TP. Research electronic data capture (REDCap). *J Med Lib Assoc* 2018 Jan 12;106(1):142-144. [doi: [10.5195/jmla.2018.319](https://doi.org/10.5195/jmla.2018.319)]
22. Wyner Z, Dublin S, Chambers C, Deval S, Herzig-Marx C, Rao S, et al. The FDA MyStudies app: a reusable platform for distributed clinical trials and real-world evidence studies. *JAMIA Open* 2020 Dec;3(4):500-505 [FREE Full text] [doi: [10.1093/jamiaopen/ooaa061](https://doi.org/10.1093/jamiaopen/ooaa061)] [Medline: [33623887](https://pubmed.ncbi.nlm.nih.gov/33623887/)]
23. FDA - SHAPE. GitHub. 2022. URL: <https://github.com/FDA/SHAPE> [accessed 2022-03-17]
24. Cross-Platform Mobile App Development. Ionic Framework. 2022. URL: <https://ionicframework.com> [accessed 2022-03-17]
25. Rare Diseases: Natural History Studies for Drug Development. U.S. Food & Drug Administration. 2019 Mar. URL: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-natural-history-studies-drug-development> [accessed 2022-03-17]
26. The Natural History Of Metachromatic Leukodystrophy Study. National Organization for Rare Diseases. 2022. URL: <https://rarediseases.org/mld-home-study/> [accessed 2022-03-17]
27. About HOME Study. HOME Study. 2022. URL: <https://mldhomestudy.iarrare.org/Home/About> [accessed 2022-03-17]
28. Moore S, Tassé AM, Thorogood A, Winship I, Zawati M, Doerr M. Consent processes for mobile app mediated research: systematic review. *JMIR Mhealth Uhealth* 2017 Aug 30;5(8):e126 [FREE Full text] [doi: [10.2196/mhealth.7014](https://doi.org/10.2196/mhealth.7014)] [Medline: [28855147](https://pubmed.ncbi.nlm.nih.gov/28855147/)]

## Abbreviations

**API:** application programming interface  
**CBER:** Center for Biologics Evaluation and Research  
**EHR:** electronic health record  
**FDA:** Food and Drug Administration  
**FHIR®:** Fast Healthcare Interoperability Resources  
**NORD:** National Organization for Rare Disorders  
**REDCap:** Research Electronic Data Capture  
**SHAPE:** Survey of Health and Patient Experience

*Edited by A Mavragani; submitted 07.04.22; peer-reviewed by L Garcia-Gancedo, A Teles; comments to author 06.07.22; revised version received 22.07.22; accepted 17.08.22; published 14.09.22.*

### *Please cite as:*

Huang H, Aschettino S, Lari N, Lee TH, Rosenberg SS, Ng X, Muthuri S, Bakshi A, Bishop K, Ezzeldin H  
*A Versatile and Scalable Platform That Streamlines Data Collection for Patient-Centered Studies: Usability and Feasibility Study*  
*JMIR Form Res* 2022;6(9):e38579  
URL: <https://formative.jmir.org/2022/9/e38579>  
doi: [10.2196/38579](https://doi.org/10.2196/38579)  
PMID: [36103218](https://pubmed.ncbi.nlm.nih.gov/36103218/)

©Haley Huang, Sofia Aschettino, Nasim Lari, Ting-Hsuan Lee, Sarah Stothers Rosenberg, Xinyi Ng, Stella Muthuri, Anirudh Bakshi, Korrin Bishop, Hussein Ezzeldin. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 14.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Implementation of a Work-Related Asthma Screening Questionnaire in Clinical Settings: Multimethods Study

Madison MacKinnon<sup>1,2</sup>, MSc; Max Moloney<sup>1,2</sup>, MSc; Emma Bullock<sup>1,2</sup>, BSc; Alison Morra<sup>1,2</sup>; Teresa To<sup>3</sup>, PhD; Catherine Lemiere<sup>4,5</sup>, MSc, MD; M Diane Loughheed<sup>1,2</sup>, MSc, MD

<sup>1</sup>Asthma Research Unit, Kingston Health Sciences Centre, Kingston, ON, Canada

<sup>2</sup>Division of Respiriology, Department of Medicine, Queen's University, Kingston, ON, Canada

<sup>3</sup>The Hospital for Sick Children, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

<sup>4</sup>Department of Chest Medicine, Hôpital du Sacré-Cœur de Montréal, Montreal, ON, Canada

<sup>5</sup>Faculty of Medicine, University of Montreal, Montreal, ON, Canada

**Corresponding Author:**

Madison MacKinnon, MSc

Asthma Research Unit

Kingston Health Sciences Centre

72 Stuart St

Kingston, ON, K7L2V7

Canada

Phone: 1 647 808 5346

Email: [13mm190@queensu.ca](mailto:13mm190@queensu.ca)

## Abstract

**Background:** A work-related asthma (WRA) screening questionnaire is currently being validated for implementation in clinical settings. To minimize barriers to integrating tools into clinical practice, a discussion of strategies for the implementation of the questionnaire has begun.

**Objective:** This study aimed to understand the benefits, feasibility, barriers, and limitations of implementing the Work-related Asthma Screening Questionnaire–Long version (WRASQ[L]) and asthma e-tools in clinical settings and propose dissemination and implementation strategies for the WRASQ(L).

**Methods:** This study was conducted in Kingston, Ontario, Canada, from September 2019 to August 2021. A workshop and 2 questionnaires were used to understand the benefits of and barriers to implementing the questionnaire in clinical settings. An expert advisory committee was established to develop the implementation and dissemination strategies. Workshops were semistructured and used thematic qualitative analysis to identify themes that provided an understanding of the benefits and limitations of and barriers to using the WRASQ(L), and e-tools in general, in clinical settings. Workshop participants included patients and health care providers, including physicians, nurses, and asthma educators, who were implementation specialists and expert electronic medical record users. A questionnaire focusing on providers' knowledge and awareness of WRA and another focusing on WRASQ(L) feedback was administered at the workshops. Advisory committee members from relevant stakeholders met 3 times to strategize implementation opportunities.

**Results:** A total of 6 themes were identified in the workshop: involving and addressing patient needs, novel data collection, knowledge translation, time considerations, functional and practical barriers, and human limitations. Questionnaire responses yielded positive feedback on the utility of the WRASQ(L) in clinical settings. All participants agreed that it is an easy way of collecting information on occupational and exposure history and could prompt a discussion between the health care provider and patient on how the workplace and exposures could affect one's asthma, increase awareness of WRA in patients and providers, and increase awareness of exposures in the workplace. Implementation and dissemination strategies were generated with input from the advisory committee.

**Conclusions:** Stakeholders and workshop participants consider the WRASQ(L) to be a useful tool that satisfies many provider needs in their clinical settings. Once validated, dissemination strategies will include developing educational materials that include the WRASQ(L), linking the questionnaire to stakeholder websites or e-toolkits, translation into other languages, leveraging health care and research networks, conference presentations, and peer-reviewed publications. Implementation strategies will include integration into electronic medical records; designing multifaceted interventions; and targeting nontraditional settings such as

workplaces, pharmacies, and research settings. The WRASQ(L) addresses many benefits of and barriers to implementation, as identified in the workshop themes. These themes will guide future implementation and dissemination strategies, noting that human limitations identified in providers and patients will need to be overcome for successful implementation.

(JMIR Form Res 2022;6(9):e37503) doi:[10.2196/37503](https://doi.org/10.2196/37503)

## KEYWORDS

work-related asthma; asthma; dissemination; implementation; e-tools; barriers; limitations; electronic medical records; EMRs; knowledge translation; mobile phone

## Introduction

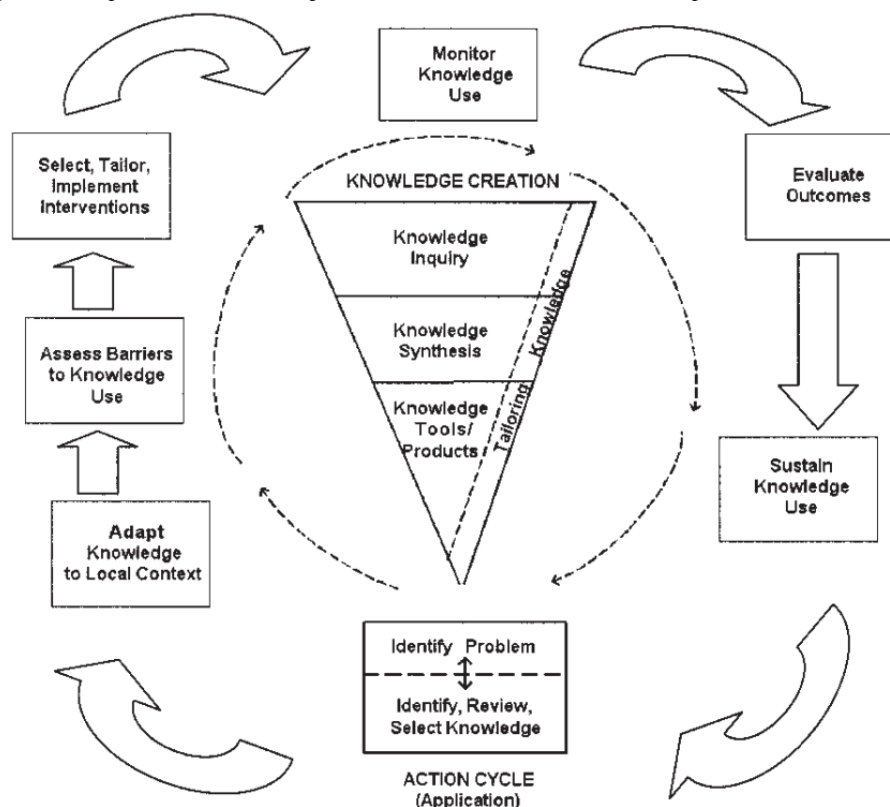
### Background of Work-Related Asthma and Knowledge Translation

Work-related asthma (WRA) is identified as asthma that is exacerbated or caused by workplace exposure and is estimated to affect as many as 25% of adults with asthma [1]. The most effective way of diagnosing WRA is through a detailed occupational history and objective measurement of lung function [2,3]. However, physicians, particularly at the primary care level, often do not have the time and resources to take a detailed occupational history, and objective measures are expensive, time consuming, and often only available in specialized centers [3,4]. This gap in screening and awareness of WRA is believed to contribute to an average of 4 years of delay between symptom

onset and diagnosis, which is associated with increased morbidity [4,5]. The diagnosis of WRA has been shown to improve patient outcomes, including health service use [6].

The knowledge-to-action (KTA) framework created by Graham et al [7] (Figure 1) outlines the elements involved in the KTA process to facilitate the implementation of research findings. The KTA framework is split into 2 dynamic concepts, called knowledge creation and knowledge action, each with its own respective phases [7]. Knowledge creation identifies the different types of knowledge and research available, and the knowledge action cycle identifies the pathway and steps to implementation, which, briefly, includes identifying a problem, adapting local knowledge to a particular context, assessing barriers, implementation, monitoring and evaluating interventions, and sustaining knowledge use [7].

**Figure 1.** The knowledge-to-action process created and reproduced from Graham et al [7], which is published under CC-BY-SA license.



In an effort to improve the accessibility, quality, and efficiency of the Canadian health care system, the Government of Canada has invested in eHealth [8]. eHealth describes the use of information and communication technology in health care and includes a wide range of technologies, including electronic patient records, telemedicine, chronic disease-monitoring

systems and management tools, electronic prescribing, and decision support tools [8-10]. In general, eHealth tools have been found to reduce symptoms, improve self-management, improve patient-provider communication, and improve overall clinical outcomes [11]. Barriers to the implementation of health tools include poor accessibility, conflicts with a practice or the



practice setting, financial incentives, individual beliefs and characteristics of health care professionals, and patient factors [12].

Integrated knowledge translation is defined as the knowledge exchange and collaboration between researchers and end users of research (eg, providers, patients, and policy makers) [13,14]. This process sees end users as partners in the research study and ensures that the asked research questions are of importance to end users rather than researchers [13,14]. End user involvement can vary widely among studies. A review examined the involvement of users in designing and evaluating self-monitoring applications for bipolar disorder. Across these studies, end user involvement ranged from just the evaluation stage of the application to all stages of research such as the evaluation, prototype design, and the concept of the application generation phases [15]. The involvement of stakeholders in the design process of interventions has been considered the “holy grail” for improvement and has been found to develop the capacity of researchers and decision-makers to engage in integrated knowledge translation processes and enhance the value of research for decision-makers [16,17]. However, few studies have included integrated knowledge translation strategies in the health care sector, and there are no clear guidelines or methodologies for end user involvement in research [16,17].

Current implementation and knowledge translation strategies for WRA, and asthma in general, have focused on the management and education of patients after their diagnosis [4]. These include prevention programs in high-risk industries [18-20], tools for patients to self-report asthma symptoms [21], and self-management plans in the form of digital applications or electronic books [22-24]. Many tools for clinicians have focused on disseminating guidelines into clinical practice to help with asthma management [4]. These include asthma care maps at the primary care level [25] and asthma care pathways in the emergency department [26]. There is a paucity of strategies or tools that focus on increasing awareness of potential WRA in individuals with asthma and few strategies or tools that are targeted to providers to increase awareness and screening of WRA [4].

## Background of the Work-Related Asthma Screening Questionnaire

The Work-related Asthma Screening Questionnaire–Long version (WRASQ(L)) was designed for implementation in clinical settings, particularly primary care, as a way of increasing awareness of WRA and screening for suspected WRA cases [27,28]. The WRASQ(L) collects occupational and exposure history, information on workplace-asthma symptom relationships, and personal protective equipment use. It includes an interpretation guide to prompt the provider on what steps to take if the patient screens positive for suspected WRA. The WRASQ(L) is available as a paper PDF, and it can be accessed through a patient or provider web-based portal via a kiosk or tablet in the waiting room or through a fillable PDF file linked to our hospital’s electronic medical record (EMR).

During its development, the WRASQ(L) was found to have good content and face validity, good test-retest reliability, and low respondent burden [28]. Although its final validation is

underway, our research team has begun to strategize the implementation and dissemination opportunities for the questionnaire. We also aimed to understand the perspectives of providers and patients on the current gaps in asthma and WRA screening and management and their perspectives on the implementation of e-tools, particularly the WRASQ(L), into clinical settings.

## Purpose and Objectives

The purpose of this paper was to report the dissemination and implementation strategies we brainstormed to maximize the impact of the research findings of the WRASQ(L)’s final validation. Using workshops, questionnaires, and an expert advisory committee, we aimed to understand the benefits, feasibility, and limitations of and barriers to implementing the WRASQ(L), and asthma e-tools in general, in clinical settings, as seen by both patients and providers, to provide valuable insights into the most effective and efficient way of implementing the screening questionnaire.

## Methods

This study was conducted in Kingston, Ontario, Canada, from September 2019 to August 2021. The participants in the project advisory committee (PAC) were from the provinces of Ontario and Manitoba. The workshop participants were all from Ontario.

## PAC Role and Engagement

Although the final validation of the WRASQ(L) is underway, a PAC was formed to oversee the questionnaire’s final validation and strategize how to implement and disseminate the questionnaire once validated. The committee was engaged at the outset, during, and at the conclusion of the validation of the questionnaire to identify the integrated and end-of-grant strategies. Stakeholders from relevant groups were invited to participate, including but not limited to Health Canada; compensation boards; professional societies; provincial and national lung, asthma, and allergy associations; public agencies; nonprofit organizations; and research groups. Potential stakeholders were invited to join the committee via email. Terms of reference were developed, and the committee met approximately biannually from 2019 to 2021 for 1 hour. The first and seventh authors cochaired the committee and led the meetings. A total of 14 members joined, including physicians, researchers, asthma educators, and nurses, from Ontario and Manitoba. A project update on the validation of the questionnaire and current knowledge translation initiatives was presented at the beginning of each meeting, after which it was opened to discussion among members. Each member was given an opportunity to provide feedback on current knowledge translation initiatives and suggestions for other end-of-grant dissemination or implementation strategies. Meeting minutes were recorded and sent to members after each meeting. Summaries of the meetings and strategies are presented.

## Asthma e-Tools Workshop and Questionnaires

We conducted 2 web-based workshops to understand health care providers’ and patients’ perspectives on the benefits and limitations of and barriers to using e-tools, including the WRASQ(L), in clinical settings. The first workshop aimed to

understand how to best integrate an asthma surveillance system, asthma indicators, and clinical guidelines in general into primary care EMRs. The second workshop focused on asthma e-tools developed by the Asthma Research Unit (ARU), including the WRASQ(L), with the aim of understanding the benefits and limitations of and barriers to using these e-tools in clinical settings. During the second workshop, there were presentations and demonstrations of the ARU's e-tools, and participants were familiarized with WRASQ(L)'s purpose, formats, and goals for implementation.

Each workshop lasted 2 hours and was conducted via Microsoft Teams. Workshops followed a semistructured question guide ([Multimedia Appendices 1 and 2](#)). The question guide was separated into different sections, each with its own individual topic or topics of discussion. Each section was allotted a specific amount of time during the 2 hours to ensure that all topics were discussed. Workshops were led by a skilled moderator—a family physician from OntarioMD who was not an asthma expert. Each participant was encouraged to speak, and the moderator moved on to the next question only once the participants had nothing more to contribute. A notetaker was also present, and the entire workshop was recorded.

The first workshop included 7 attendees. A total of 6 attendees were selected by the organization (OntarioMD) that facilitated the workshop because of their expertise in EMR use and implementation. Of the 6 attendees, 5 (83%) attendees were family physicians, and 1 (17%) was a nurse practitioner. Our research team recommended that another family physician with a special interest in respiratory health participate as a guest expert. The second workshop contained 6 participants, of whom 4 (67%) were selected by OntarioMD, 2 (33%) were family physicians, 1 (17%) was a nurse practitioner, and 1 (17%) was a patient. Our team recommended that the same family physician and nurse practitioner who was also an asthma educator participate. The practitioners were all based in Ontario and had a wide range of locations of practice, from rural to urban centers. The guest participants added by our research team were purposeful and had extensive knowledge of asthma e-tools and EMRs in Ontario. Some participants knew the researchers and were familiar with previous work, whereas others were unfamiliar.

Questionnaires were sent to the participants before and after the second workshop ([Multimedia Appendices 3 and 4](#)). The first questionnaire, which was sent before the workshop, aimed to assess the providers' knowledge and awareness of WRA. Providers were first asked whether they discussed occupational history with patients with suspected or confirmed asthma. If so, providers were asked whether this information was recorded (in EMRs, paper charts, or not at all). If this information was not recorded, providers were asked to provide a reason for not recording it. Subsequently, providers were asked whether they discussed potential WRA, workplace exposures, and the potential relationship between workplace exposures and asthma symptoms in the workplace with their patients. The second questionnaire, which was sent after the workshop, asked for specific feedback on the WRASQ(L). One of the questions used a Likert scale to understand how much participants agreed with statements about the WRASQ(L)'s utility in prompting a

discussion on the relationship between workplace exposures and asthma, raising awareness of WRA and potentially harmful exposures, collecting occupational history, improving screening, and increasing referral time to a specialist for WRA. We then asked whether providers would consider administering the WRASQ(L) in their practice and, if so, asked for which purpose (screening for WRA, collecting occupational history, collecting information about the relationship between the patient's asthma symptoms and workplace exposures, initiate a conversation about the topic of workplace-symptom relationship with patients, or others) and in what format.

The questionnaire responses were tallied by frequency and percentage. A total of 3 team members from the ARU participated in the thematic qualitative analysis of the workshops [29]. All the team members engaged in reflexivity throughout the research process [30]. The team members met at the beginning of the data analysis to discuss and record how their personal experiences and biases could influence their interpretation of the results. All were a part of the ARU and familiar with the e-tools, literature related to WRA diagnosis, and reporting and implementation of asthma e-tools. One of the team members took notes at all research team meetings, and all members recorded notes in a memo throughout the analysis.

The workshop audio was transcribed verbatim and rechecked by a different team member. All team members reviewed the transcripts multiple times, rewatched the recorded workshops, and reviewed their notes taken during the workshops to familiarize themselves with and obtain a general and descriptive sense of the data.

All members were engaged in the coding of the data. Coding was conducted separately, followed by regular team meetings to discuss codes, overarching themes, and impressions from the data. Codes were added to a codebook that was refined and narrowed down with subsequent coding sessions. Transcripts were reviewed and recorded until they agreed that data saturation had been met.

Relationships between codes were identified using tables and mind maps to organize the codes into overarching preliminary themes. Themes were reviewed over multiple group meetings with input from all team members and refined until the key themes that defined the essence of the data were agreed upon. The key themes were named and defined, and quotations that provided sufficient evidence of the themes were selected. Any disagreement during the coding or identification of themes was discussed and resolved by the group.

### Ethics Approval

The study was reviewed for ethical compliance by the Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (approval numbers: TRAQ# 6029444 and 6019013).

## Results

### PAC Strategies

A total of 2 members affiliated with the Workplace Safety and Insurance Board Champions Program, a program working to

implement occupational health modules in Ontario medical schools, suggested discussing the WRASQ(L) with a Queen’s University representative. There was also a discussion of partnering with the Lung Health Foundation as they are creating an e-module for providers on WRA. The PAC noted that translation into other languages should be considered, particularly in Chinese, as asthma is prevalent in the Asian community. They also suggested using the WRASQ(L) as a validated tool for research and as a way of placing occupational information that is clinically useful and relevant into EMRs, as 1 member mentioned that the Ministry of Labor was working on such an initiative.

One of the members noted that the European Respiratory Society Task Force was examining validated questionnaires to be used clinically and in research for WRA surveillance. Other health care provider networks were suggested to be leveraged, such as patient advocacy through the Lung Health Foundation and certified respiratory educators through the Canadian Network for Respiratory Care. A final suggestion was to consider implementing the WRASQ(L) in pharmacy settings for patients with suspected WRA who are yet to see a physician.

**Questionnaires**

The questionnaire asking providers about their awareness of WRA was provided to those participants of the second workshop who were health care practitioners. Therefore, it was filled out by 83% (5/6) of participants. The questionnaire had a response rate of 100%. All participants (5/5, 100%) said they discussed their occupational history with patients with suspected or confirmed asthma, and most (4/5, 80%) stated that they recorded their occupational history in their EMR. One of the participants said that they asked the patient whether they wanted the detailed work history recorded or whether they just wanted an overview of it but did not specify where they placed it. Approximately 80% (4/5) of participants reported that they routinely discussed

the potential relationship between workplace and asthma symptoms with patients with suspected or confirmed asthma. Approximately 20% (1/5) of participants said whether they are going to discuss depends on the age and stage of their asthma. All participants (5/5, 100%) reported inquiring about the exposures with which patients were in contact at their workplace in those with suspected and confirmed asthma. Finally, all participants except 20% (1/5) said that they discussed the management of asthma in relation to the workplace with patients with confirmed or suspected asthma.

The questionnaire that asked for WRASQ(L) feedback had a response rate of 80%. Overall, participants felt it was beneficial and could prompt a discussion between the health care provider and patient on how the workplace and exposures could affect one’s asthma, increase awareness of WRA in patients and providers, and increase awareness of exposures in the workplace. All participants strongly agreed that it was an easy way of collecting information on occupational and exposure history. Approximately 75% (3/4) of participants strongly agreed that the WRASQ(L) could improve the screening of WRA at the primary care level, speed the time to referral to a specialist, and decrease the time to diagnosis.

**Asthma e-Tool Workshops**

*Overview*

A total of 6 themes explained health care provider preferences regarding the use of e-tools, particularly the WRASQ(L), in clinical settings, with subthemes that organize the narrative. These themes can be categorized into 3 benefits, 2 key barriers or limitations, and 1 considered both a benefit and a barrier or limitation. The themes were as follows: (1) involve and address patient needs, (2) novel data collection, (3) knowledge translation, (4) time considerations, (5) functional or practical barriers, and (6) human limitations (Table 1).

**Table 1.** Brief description of themes.

Theme	Benefit or barrier	Description
Involve and address patient needs	Benefit	It is important for patients to feel involved in their care, and thus, tools should enable this. This can be done by having flexibility regarding when the tool can be used and the format of the tool by providing feedback to patients and considering their fatigue and fears.
Novel data collection	Benefit	Tools should fill a gap in data collection or provide a unique way of collecting data.
Knowledge translation	Benefit	Tools are beneficial when they translate knowledge from the specialist to the primary care provider or the provider to the patient.
Time considerations	Benefit and barrier	Any tool that saves the provider time is incredibly beneficial; however, if it takes too much time to learn, use, and implement, it is a barrier.
Functional and practical barriers	Barrier	Limitations in technology, particularly seamless integration of tools into electronic medical records and for patient use, and resources will impede the use of the tool in practice.
Human limitations	Barrier	Provider and patient attitudes and behaviors, such as mistrust of tools and personal biases and fears, and the tendency to not reuse tools by stakeholders are all human limitations to the uptake of tools.

**Involve and Address Patient Needs**

The participants noted that e-tools that involve or inform patients about their care are beneficial to clinical practice. Patients take pride in and give importance to being involved in their care and seeing themselves as “partners” with providers in their care.

Tools are not considered useful to the patient population if there is “no clear follow-up,” and a feedback loop to inform patients is beneficial:

*I think the idea of the feedback loop is [a] really important one...Because many times we just collect*

*data and we don't actually let you know, or to what end, or to let you give any sort of response on to what we've done as a result of that. So, I think that there's potential for the applications or these tools to do that in real-time.*

Flexibility in when patients use e-tools or the format in which they are administered has been frequently discussed. Questionnaire fatigue was mentioned, with concern over how fatigue can affect the authenticity or accuracy of the answers. Another concern was the mistrusted answers from patients who complete questionnaires in stressful situations as they are just trying to get the questionnaire “out of the way.”

In discussing potential solutions, a popular option from both patient and physician peer leaders was flexibility, both in the timing and format of the questionnaire or e-tool administration. The options discussed were before visits, in the waiting room, or during the visit while the provider was doing other clinical activities, ultimately wherever made the patient most comfortable:

*...in the waiting room I would love to have something to fill out...It is the perfect opportunity. I think the way I would prefer it to happen would [be] to get an email a couple days before an appointment and have the opportunity to fill out but if I don't, then I'm handed a tablet at the appointment visit to fill it out right. I think you gotta use both strategies, not one or the other.*

### **Novel Data Collection**

Tools must fill a certain gap in data collection or provide a unique aspect to data collection to be considered beneficial by providers. In the workshops, participants mentioned the underreporting and undermanagement of asthma in the population. For example, they found it difficult to document occupational history in the EMRs. A tool that fills these gaps and provides an opportunity for these data to be recorded would provide a major benefit to clinical practice.

Tools that present data in a unique manner, such as through visualizations of the data, benefit practice. The asthma educator participant emphasized that tools that can show novel trends in the data for patients are very beneficial:

*I've had patients say it is really helpful to see how that tool is able to give me a visual on how this has improved my life...*

### **Knowledge Translation**

Tools must facilitate the translation of knowledge from specialists to generalists or from providers to patients. Participants noted this comes from a clear understanding of what the tool does and how to use it and, ultimately, how to use it to improve their care:

*I think that part of it needs to be solved just in...the knowledge translation, what is this tool actually for*

The providers discussed that the integration of the tools into clinical settings is an incredibly important step in the knowledge translation process. It facilitates the movement of information

to the provider or patient and helps physicians and patients “manage their issues in the most optimal way.”

### **Time Considerations**

Time was a central theme in both workshops, particularly for providers. Participants emphasized that tools with time-saving features were incredibly helpful and more likely to be used. Automated features, such as drop-down menus or a proactive reminder to use the tool, were viewed favorably by the participants. Participants emphasized that tools should be efficient and easy to use so that they do not affect their practice:

*We have to make it as easy as possible; I think that's kind of the key...otherwise people are not going to do it.*

Conversely, if a tool takes too much time to learn or use, then it is a major limitation or barrier to using the tool. One of the participants noted that time constraints in the clinic could prevent them from using the tool, even if it was already implemented.

### **Practical and Functional Limitations**

One of the main limitations discussed by the participants was practical or functional limitations, in other words, a lack of resources and technological limitations in accessing, using, and implementing e-tools. Almost all participants emphasized that tools need to be seamlessly linked or integrated into the EMR. A seamless linkage in practice and in using e-tools is considered when the provider does not have to leave their EMR or patient charts to access e-tools or other programs. It also includes the transfer of data from the tool or program back to patient charts. Many participants said that having to leave their EMR to use a tool is a major barrier, and the interconnectedness of tools is currently lacking:

*I meant the integration itself is just so important...I mean, beyond just leaving the environment you're in, which would be a real pain and is certainly a barrier to adopting these things, but like the integration piece allows you to bring data in...But nobody's really figured this out or I don't know of anybody that's figured this out...*

Providers frequently noted that a lack of physical, human, technological, and financial resources are major barriers to implementation. Managing the data that comes from these tools was noted to be difficult if there were not as many resources available, causing more work for the provider, which may ultimately stop providers from using the tool.

### **Human Limitations**

#### **Provider Behavior**

Providers' behaviors and preconceived ideas or biases are major limitations to the use and implementation of e-tools. Fear and apprehension about how e-tools could negatively affect providers were mentioned many times:

*I think on the other side there's always this fear that the data is somehow going to be used for, you know, negotiations or if it's in the wrong hands is going to*

*be used against the physician in some sort of way which, which is obviously far from the truth.*

Participants corroborated this from their own experiences when they implemented their own e-tools in the form of a dashboard. They noted that stakeholders asked how the information would be used to “punish” them, and there was a “suspicion” they were sharing the data.

Convincing providers to try a new tool was another barrier. It was considered difficult to market a new tool to providers, and providers were noted as having a “defeatist attitude” that new e-tools are “not useful at all” when they do not work perfectly, meet the providers’ expectations immediately, or are slow to be implemented. Behavior change is required for providers to adopt and implement a new tool. Incentives, such as funding, and quick turnaround of information have been mentioned as ways of inciting behavior change.

### Trust and Proof of Value

Trust must be established between the tool and the patient or provider. The user needs to feel that they can trust the tool and the information provided and that it will make a difference in their practice. Trust was established by determining whether the provided data were accurate. One of the participants mentioned that when discussing new e-tools or implementations with providers, “the immediate discussion goes to ‘well that’s not accurate.’” Providers prioritize and need to see clear and accurate data to adopt a new tool as this is proof that the tool will be valuable to them. If providers and patients do not see the value of the tool, then it is unlikely to be reused. Both provider and patient participants expressed that tools or apps can be forgotten if not deemed useful:

*I use a lot of apps in my practice on my mobile devices and the good ones I use regularly and the ones that aren’t that great you stop using.*

If patients and providers can see how it improves practice, it will establish a level of trust and change behavior. Seeing and understanding the proof of value of the tool can lead to the desired outcome of sustained use of the tool in practice. Overcoming alert fatigue, which was mentioned twice by participants, and “rewarding good outcomes and good behaviour” are the means to achieve this outcome.

## Discussion

### Principal Findings

We aimed to understand the benefits, feasibility, limitations of and barriers to implementing asthma e-tools in general and, specifically, the WRASQ(L), in clinical settings, as seen by both patients and providers. Through the focus groups and PAC, we gained information from both end users and specialists in the field. Our findings provide insights into the potential implementation and dissemination opportunities for WRASQ(L). Our findings suggest that the questionnaire, and e-tools in general, are considered useful in clinical settings and have the

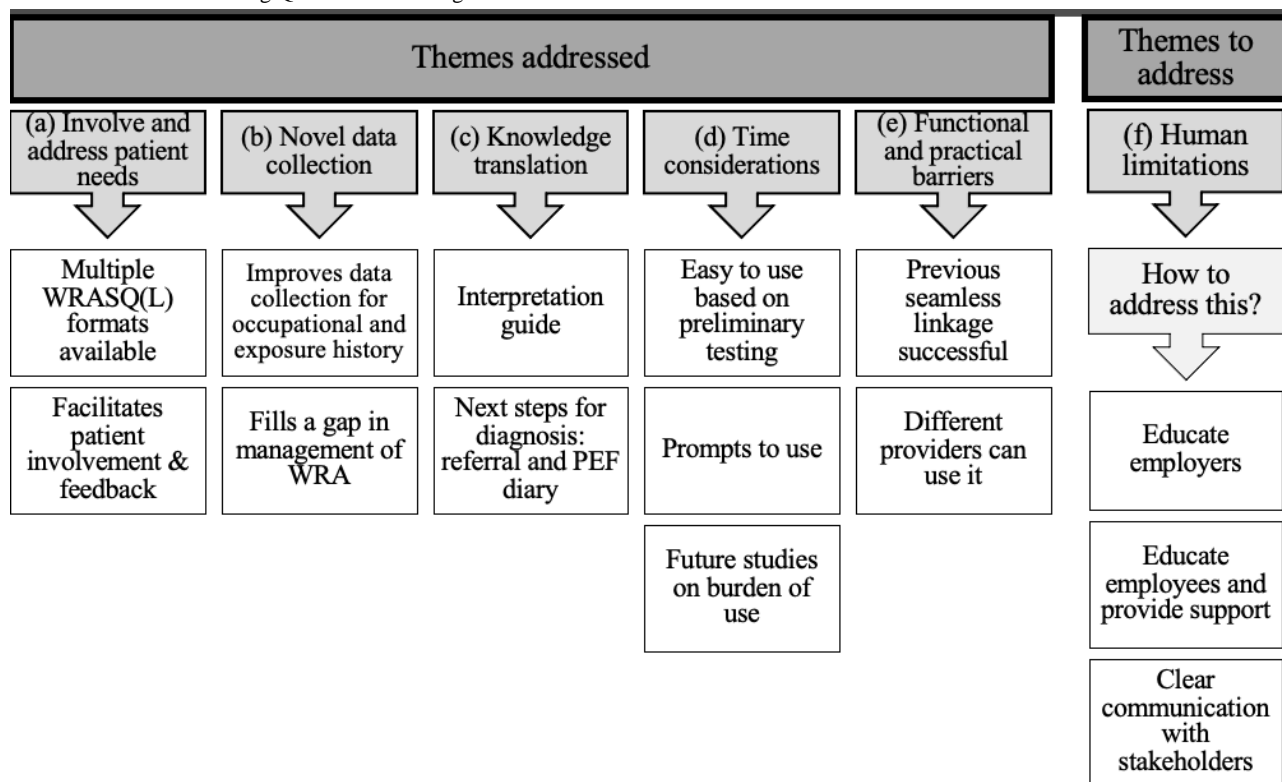
potential to greatly improve practice. The identified key barriers need to be overcome to facilitate the adoption of asthma e-tools.

Through the preworkshop survey, we found that workshop participants reported discussing relevant information with their patients who had suspected and confirmed asthma, such as symptom-workplace relationships and management and recording occupational history. These results contrast with those found in the literature, which identifies a major gap in care in primary care settings for WRA, especially in taking detailed occupational history, workplace exposure history, or discussion on how asthma might be work related [4,31]. Our results may differ because of the selection bias of the participants. Some participants had a keen interest in respiratory health and were expert users of EMRs; hence, they may ask for and record this information in their practice. The postworkshop questionnaire results were encouraging; all participants agreed that the WRASQ(L)’s implementation would be beneficial, that it could speed up the time to referrals to a specialist to ultimately decrease the delays in diagnosis, and that it was a useful tool to collect occupational history.

Our workshop concentrated on the benefits and limitations of and barriers to implementing e-tools in clinical practice. We used an inductive approach to understand the health care provider and patient perspectives on e-tools [29]. The themes pertained to involving patients in their care, creating a new type of data collection, facilitating knowledge translation, saving and not taking up too much time, and overcoming functional barriers and human limitations. The findings suggest that, overall, e-tools are considered beneficial in clinical settings but only if their implementation and use can overcome the identified barriers. The findings also provide important context and knowledge on how to best implement WRASQ(L) and ensure the future use of the tool.

Although many of the themes identified in our workshops have been reported in the literature, the WRASQ(L) addresses and expands on these themes in the context of implementing a WRA tool in primary care EMRs (Figure 2). The workshop results also established a novel theme of knowledge translation. It was established that patients take pride in their care, want to be involved in their care, and want feedback. This has been noted in the literature, particularly in studies involving patients with chronic diseases [32,33]. A benefit of the WRASQ(L) is its ability to easily collect occupational history and its many different forms. It can be filled out on paper, electronically, or via a kiosk that is accessed from a portal via a smartphone or tablet. This allows flexibility in how patients complete the WRASQ(L). If completed in the clinic, the questionnaire can be interpreted immediately, which would provide immediate feedback to the patient. These benefits address the themes of “Involve and Address Patient Needs” and “Novel Data Collection.” The WRASQ(L) addresses questionnaire fatigue as well, as it has a low respondent burden and takes, on average, <10 minutes to complete (mean 7.2, SD 3.8 minutes), which makes it a timely questionnaire to complete [28].

**Figure 2.** Themes addressed and themes to address by the WRASQ(L). PEF: peak expiratory flow; WRA: work-related asthma; WRASQ(L): Work-Related Asthma Screening Questionnaire–Long version.



Time constraints have been stated as the most important barrier to taking an occupational history; therefore, this theme emulates one of the key barriers to diagnosing WRA [4,31]. Our participants showed enthusiasm for automatic features, and it has been found that successful implementation strategies previously included the use of reminders for the tool [12]. Although the WRASQ(L) does not have reminders, it has easy-to-fill options and a prompt to fill out the questionnaire in the asthma management systems to which it is seamlessly linked. The literature, and workshop findings, suggest that implementation should occur within a realistic time frame and that the clinical utility of tools is maximized when the tool is time efficient and easy to use [12,34]. The WRASQ(L) has been found to have a low respondent burden and good test-retest ability; therefore, it has been deemed easy to administer. No studies have examined the burden of using the questionnaire in clinics; however, this leads to future studies after implementation. Our participants greatly emphasized the vertical integration of tools into EMRs. The WRASQ(L) has already been successfully integrated into an asthma management system that seamlessly links to a fillable PDF file and kiosk version.

An interesting result was that providers and patients themselves could be barriers to adoption. Providers’ preconceived fears, notions, and attitudes regarding the implementation and use of new tools were very evident. Behavior, or more specifically, attitude change, is very important for implementation as it is unlikely that the tool will be implemented if the provider is not receptive to change or their concerns are not addressed [12]. Therefore, although human limitations are a general barrier to implementation reported in the literature, researchers should identify the specific human limitations that relate to the conditions they are studying, which could affect their

implementation of the e-tool. Behaviors by providers and participants specific to occupational diseases are foreseeable barriers to implementing our questionnaire. Providers hesitate to diagnose and manage occupational diseases because of the burden of submitting a compensation claim [35,36]. Patients avoid discussing their health concerns in the workplace for fear of the stigma associated with injured workers; in particular, fear that coworkers or managers will think they are abusing the system or malingering [35,37]. There is a hesitancy to file claims or report health issues, as well as the fear of losing their jobs [4,35,38]. Thus, we need to address these specific concerns regarding human behaviors in work-related conditions when moving forward with our implementation. In addition, this theme showed that clear communication about the purpose of the tool is needed, and providers need to feel that they understand and trust the tool. This is established by longevity; if patients or providers continually use the tool, then this is the desired outcome. Ultimately, sustained use of the tool in clinical settings is the overall goal.

A novel theme identified by our focus groups was that it would be ideal if a tool could be a conduit for knowledge translation; that is, tools were considered beneficial when they provided knowledge exchange between providers and from providers to patients. This is particularly beneficial for an underreported disease such as WRA, and we believe that the WRASQ(L) is able to address this. The WRASQ(L)’s interpretation guide outlines the recommended steps in the care of the patient; thus, it conducts knowledge translation by bringing this expert knowledge from specialists to the primary care level and the patients. In addition, by simply using the WRASQ(L) once, patients and providers are made aware of the potential harmful exposures that could cause WRA, the relationship between

asthma symptoms and the workplace that could indicate WRA, and the importance of taking an occupational history. Thus, the creation and implementation of e-tools should prioritize an element of knowledge transfer between users and the literature for ideal and long-term implementation.

Our workshops identified how the WRASQ(L) could benefit clinical practice from the viewpoint of end users. Once

limitations are addressed, the strategies from the specialists in the PAC can be used to implement the questionnaire in the field. Discussions from our PAC focused on leveraging and connecting with other health care networks and stakeholders with whom many members were affiliated or had worked with before. Ultimately, we proposed several strategies for use once the final validation of the WRASQ(L) was completed ([Textbox 1](#)).

**Textbox 1.** Summary of implementation and dissemination strategies.

<p><b>Implementation strategies</b></p> <ul style="list-style-type: none"> <li>• Integrate into electronic medical records <ul style="list-style-type: none"> <li>• OntarioMD's dashboard</li> <li>• Collaborate with electronic medical record vendors, Ministry of Labor, and Ministry of Health</li> </ul> </li> <li>• Design multifaceted interventions <ul style="list-style-type: none"> <li>• Prompts and reminders</li> <li>• Performance indicators</li> <li>• Audit and feedback</li> </ul> </li> <li>• Target nontraditional settings <ul style="list-style-type: none"> <li>• Workplaces</li> <li>• Pharmacies</li> </ul> </li> <li>• Implement in research settings as well as clinical</li> </ul> <p><b>Dissemination strategies</b></p> <ul style="list-style-type: none"> <li>• Develop educational material <ul style="list-style-type: none"> <li>• Workplace Safety and Insurance Board Champions Program</li> <li>• Lung Health Foundation's e-module for providers on work-related asthma</li> </ul> </li> <li>• Link to websites or electronic toolkits</li> <li>• Lung Health Foundation's "current educational strategies" for providers</li> <li>• Canadian Thoracic Society toolkit</li> <li>• Translate to other languages</li> <li>• Leveraging existing health care and research networks: <ul style="list-style-type: none"> <li>• Certified respiratory educators: Canadian Network for Respiratory Care and Primary Care Asthma Program</li> <li>• Canadian Thoracic Society Asthma Clinical Assembly</li> <li>• American Thoracic Society and European Respiratory Society Work-Related Asthma Taskforces</li> <li>• Center for Research Expertise in Occupational Disease</li> </ul> </li> <li>• Conference presentations</li> <li>• Peer-reviewed publication</li> </ul>
---

The dissemination strategies proposed by the PAC included common actions such as conference presentations and peer-review publications; however, members also suggested implementing the WRASQ(L) as educational material and linking it to websites or e-toolkits. These strategies have the potential to have a 2-fold effect. They could not only increase the use of the WRASQ(L) by providers but could also increase awareness of WRA. This would address a major concern with

WRA by addressing the paucity of strategies or tools that focus on increasing awareness and screening for potential WRA among providers [4]. The translation of the questionnaire into other languages would further contribute to this. This would allow the administration of the questionnaire to populations at a higher risk of asthma, such as many Asian communities. Furthermore, it would increase the accessibility of the questionnaire to other providers in Canada and around the world,

thus, potentially increasing the use of the questionnaire and increasing awareness of WRA in these places. Finally, the PAC members noted that other health care networks could be included more frequently in the dissemination process. Certified respiratory educators were particularly noted as a key network to leverage as they are imperative to the education of patients. In addition, focusing efforts on leveraging other research networks such as the Centre for Research Expertise in Occupational Disease would provide the opportunity for the questionnaire to be used in other research settings (a noted implementation strategy discussed in the following paragraph) and to be seen by other researchers, which has the potential to increase the use and awareness of the questionnaire.

Implementation strategies were more general than dissemination strategies but still provided guidance on how to increase the use of the questionnaire and awareness of WRA. Targeting nontraditional settings such as workplaces, pharmacies, and research settings has the potential to increase awareness of WRA in these places when it might be lacking. For example, placing the questionnaire in a workplace that has a high risk of WRA could inform workers and employers of the potential for WRA in the setting. This would not only create awareness of the issue but could also prompt employers to be aware of their workers' conditions, mitigate risk with the provision of personal protective equipment, increase communication of the potential for WRA between employers and employees, and decrease stigma or fear of reporting WRA. Similarly, implementing the questionnaire in pharmacies and research settings would make other health care professionals such as pharmacists and other researchers more aware of WRA, despite not using the questionnaire for clinical purposes. Integration into EMRs was a concrete strategy that was proposed, and partnering with existing dashboards such as OntarioMD's Insights4Care dashboard would allow for easy implementation. As stated, one of the members suggested that the Ministry of Labor was working on a way of including clinically useful occupational information in EMRs. Approaching bodies such as this early in their implementation process would ensure the WRASQ(L) is included as well. As implementation is a lengthy process, it is wise to explore these options so that implementation and dissemination can be timely once the questionnaire is validated. To the best of our knowledge, this method of approaching end users through our workshops and experts via the PAC before the actual implementation of the questionnaire is novel. Obtaining this knowledge will not only allow the implementation to be timely but also create and guide a robust implementation strategy for the questionnaire to maximize its use.

Our findings address both the action cycle and knowledge creation concepts in the KTA framework. Each phase of the knowledge creation concept allows researchers to tailor their activities to the needs of their ideal stakeholders and customize their methods of dissemination [7]. The workshop provided valuable insights into how we can tailor the WRASQ(L) to satisfy the concerns of stakeholders (patients and providers). Both the PAC and workshop findings addressed the steps "Adapt Knowledge to Local Context" and "Assess Barriers and Knowledge Use" in the action cycle. By reviewing current KTA

knowledge and initiatives with the PAC and discussing gaps in management in the workshop, we adapted current knowledge to our context, which is the improvement of asthma management with e-tools. Many barriers were assessed by both groups. This will allow us to move confidently into the "Select, Tailor and Implement Interventions" phase to implement our concrete strategies while noting the barriers we may face.

A limitation of this study is that it does not include a theoretical framework for assessing the determinants of successful implementation of the questionnaire; however, we believe that the specific tool and context in which we aim to implement the tool benefited from an inductive approach. There is some concern that the use of frameworks can influence deductive analysis, bias researchers, and unconsciously force themes discovered into preconceived categories [39,40]. This was a potential concern because of the context of our study and our tool. WRA is a subtype of asthma that is unfamiliar to many patients and providers [4,5], and we aimed to understand the use of a tool for this specific disease if implemented in EMRs in primary care. In addition, we are at a very preliminary stage in the implementation process, and to the best of our knowledge, a similar study has not been completed previously for WRA screening tools. Thus, an open-ended approach allowed us to gather as much information as possible from stakeholders about the benefits of and barriers to implementing the tool in this context. The identified barriers and themes, along with a theoretical framework, could guide a robust and efficient implementation strategy once the validation is complete.

Our methods and findings allow the research team to approach these ideas and address potential limitations and barriers early to ensure efficient and timely implementation of WRASQ(L). These methods may be applied to other studies that validate e-tools, particularly studies that consider the involvement of end users and experts to discuss implementation strategies before the completion of the validation.

### Limitations

This study has several limitations. The small sample size and lack of use of the theoretical framework in this study may have reduced the generalizability of the results. Selection bias was present in our participants, as 33% (2/6) of participants with a keen interest in respiratory health (a family physician and an asthma educator) were invited. This decision was made to address what we felt was a serious limitation in the peer leaders selected by OntarioMD, as they lacked practical primary care expertise in asthma. A total of 3 workshop participants were familiar with the WRASQ(L) and ARU e-tools. Despite this, we felt that their contribution to the focus groups was beneficial, as they brought practical primary care asthma expertise into the discussion. Of the 3 participants, 2 (67%) had no experience using the WRASQ(L) in clinical settings, nor had they been asked whether they would use the tool in their practice. One of the participants who had used the questionnaire before in clinical settings was considered an important contributor as they could provide a unique perspective of how patients responded to the tool. All 3 participants were expert EMR users; thus, their insight was valuable. The use of a third-party moderator for the workshops mitigated bias by ensuring that all participants had



an equal chance of contributing to the discussion. Finally, only one-half of a workshop focused on the WRASQ(L), making it challenging to identify a clear overarching implementation strategy. This was offset by the PAC, whose sole purpose was to discuss the implementation and dissemination of the questionnaire; however, the PAC members were not implementation specialists. It may be beneficial to conduct another workshop for only the questionnaire with a larger sample size.

### Conclusions

By addressing both the knowledge action and knowledge creation phases in the KTA framework, we identified key strategies to support the implementation of the WRASQ(L). Participants perceived the high utility of this WRA screening questionnaire in clinical settings and that it addressed many

themes identified in our workshops relating to the implementation of e-tools in primary care EMRs. The workshop results and PAC recommendations will guide future dissemination and implementation initiatives and may be generalizable to other asthma e-tools.

Dissemination strategies will include incorporating the questionnaire in educational material, linking the questionnaire to websites or e-toolkits, translating it into other languages, and leveraging health care and research networks. Implementation strategies will include the integration of the WRASQ(L) into EMRs; designing multifaceted interventions; and targeting nontraditional settings such as workplaces, pharmacies, and research settings. The theme or barrier of human limitations may require more time and effort to overcome once the implementation of the questionnaire begins.

---

### Acknowledgments

The authors are grateful to the project advisory committee members for their valuable input: Victoria Arrandale, Occupational Cancer Research Center, Occupational and Health Environmental Division of the Dalla Lana School of Public Health; Jennifer Dyck, Occupational Health Nurse Prevention Consultant; Anne Ellis, Canadian Society of Allergy and Clinical Immunology Executive member representative; Krystelle Godbout, Quebec Heart and Lung Institute; Linn Holness, Centre for Research Expertise in Occupational Disease; Allen Kraut, Medical Director, Occupational Health at the University of Manitoba; Jennifer MacKinnon, Lung Health Foundation (Education); Jessica Schooley, Primary Care Asthma Program Educator, Kingston Health Sciences Centre Asthma Program; Clare Ramsey, Canadian Thoracic Society and Manitoba Lung Association; and Susan Tarlo, Canadian Thoracic Society. The authors would also like to acknowledge OntarioMD and the OntarioMD peer leaders for their participation in the workshops and their valuable input.

M MacKinnon received a stipend from the Worker's Compensation Board of Manitoba and Queen's University in support of her graduate studies. The Worker's Compensation Board of Manitoba provided funding to the PAC. This project was supported by the Southeastern Ontario Academic Medical Organization Innovation Fund of the Alternative Funding Plan for the Academic Health Sciences Centers of Ontario, Government of Ontario. The funding sources were not involved in the design of the study, collection and interpretation of the literature review, writing of this report, or decision to submit this study for publication.

---

### Authors' Contributions

M MacKinnon and MDL conceived the study questions and design. The workshops were organized by AM. Data collection and entry were performed by M MacKinnon, M Moloney, AM, and EB. Data analysis was led by M MacKinnon with inputs from M Moloney and EB. M MacKinnon wrote the first draft of the manuscript, with revisions by MDL, M Moloney, EB, and AM. TT and CL provided expert feedback on the draft and made revisions to the manuscript. All authors have read and approved the final manuscript.

---

### Conflicts of Interest

MDL has received grants outside the submitted work paid directly to Queen's University from the Canadian Institutes of Health Research (subgrant from Ottawa Health Research Institute), Manitoba Workers Compensation Board, Queen's University, and GlaxoSmithKline, as well as honoraria from the Canadian Thoracic Society for codevelopment and copresentation of a Severe Asthma PREP course and honoraria from AstraZeneca for participation in the Precision Program Advisory Board. CL sits on the occupational lung disease committee of the Commission des normes, de l'équité et de la santé et de la sécurité du travail.

---

#### Multimedia Appendix 1

Interview guide for workshop 1.

[\[PDF File \(Adobe PDF File\), 92 KB - formative\\_v6i9e37503\\_app1.pdf \]](#)

---

#### Multimedia Appendix 2

Interview guide for workshop 2.

[\[PDF File \(Adobe PDF File\), 95 KB - formative\\_v6i9e37503\\_app2.pdf \]](#)

## Multimedia Appendix 3

Questionnaire assessing providers' knowledge and awareness of work-related asthma.

[\[PDF File \(Adobe PDF File\), 68 KB - formative\\_v6i9e37503\\_app3.pdf\]](#)

## Multimedia Appendix 4

Questionnaire for the Work-Related Asthma Screening Questionnaire—long version feedback.

[\[PDF File \(Adobe PDF File\), 81 KB - formative\\_v6i9e37503\\_app4.pdf\]](#)

**References**

1. Tarlo SM, Balmes J, Balkissoon R, Beach J, Beckett W, Bernstein D, et al. Diagnosis and management of work-related asthma: American College Of Chest Physicians Consensus Statement. *Chest* 2008 Sep;134(3 Suppl):1S-41S. [doi: [10.1378/chest.08-0201](https://doi.org/10.1378/chest.08-0201)] [Medline: [18779187](https://pubmed.ncbi.nlm.nih.gov/18779187/)]
2. Tarlo S, Liss G, Blanc P. How to diagnose and treat work-related asthma: key messages for clinical practice from the American college of chest physicians consensus statement. *Pol Arch Med Wewn* 2009 Oct;119(10):660-666 [FREE Full text] [Medline: [19847143](https://pubmed.ncbi.nlm.nih.gov/19847143/)]
3. Lemièrè C. Diagnosing occupational asthma: insight from induced sputum. *Can J Physiol Pharmacol* 2006 Jan;84(1):1-4. [doi: [10.1139/y05-135](https://doi.org/10.1139/y05-135)] [Medline: [16845884](https://pubmed.ncbi.nlm.nih.gov/16845884/)]
4. MacKinnon M, To T, Ramsey C, Lemièrè C, Lougheed MD. Improving detection of work-related asthma: a review of gaps in awareness, reporting and knowledge translation. *Allergy Asthma Clin Immunol* 2020 Aug 06;16(1):73 [FREE Full text] [doi: [10.1186/s13223-020-00470-w](https://doi.org/10.1186/s13223-020-00470-w)] [Medline: [32922457](https://pubmed.ncbi.nlm.nih.gov/32922457/)]
5. Santos MS, Jung H, Peyrovi J, Lou W, Liss GM, Tarlo SM. Occupational asthma and work-exacerbated asthma: factors associated with time to diagnostic steps. *Chest* 2007 Jun;131(6):1768-1775. [doi: [10.1378/chest.06-2487](https://doi.org/10.1378/chest.06-2487)] [Medline: [17505048](https://pubmed.ncbi.nlm.nih.gov/17505048/)]
6. Lemièrè C, To T, de Olim C, Ribeiro M, Liss G, Lougheed MD, et al. Outcome of work-related asthma exacerbations in Quebec and Ontario. *Eur Respir J* 2015 Jan;45(1):266-268 [FREE Full text] [doi: [10.1183/09031936.00096114](https://doi.org/10.1183/09031936.00096114)] [Medline: [25359335](https://pubmed.ncbi.nlm.nih.gov/25359335/)]
7. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? *J Contin Educ Health Prof* 2006;26(1):13-24. [doi: [10.1002/chp.47](https://doi.org/10.1002/chp.47)] [Medline: [16557505](https://pubmed.ncbi.nlm.nih.gov/16557505/)]
8. eHealth. Government of Canada. 2010. URL: <https://www.canada.ca/en/health-canada/services/health-care-system/ehealth.html> [accessed 2022-07-25]
9. Health Quality Ontario. Electronic tools for health information exchange: an evidence-based analysis. *Ont Health Technol Assess Ser* 2013;13(11):1-76 [FREE Full text] [Medline: [24194799](https://pubmed.ncbi.nlm.nih.gov/24194799/)]
10. Rios-Zertuche D, Gonzalez-Marmol A, Millán-Velasco F, Schwarzbauer K, Tristao I. Implementing electronic decision-support tools to strengthen healthcare network data-driven decision-making. *Arch Public Health* 2020 Jun 18;78:33 [FREE Full text] [doi: [10.1186/s13690-020-00413-2](https://doi.org/10.1186/s13690-020-00413-2)] [Medline: [32566223](https://pubmed.ncbi.nlm.nih.gov/32566223/)]
11. Lancaster K, Abuzour A, Khaira M, Mathers A, Chan A, Bui V, et al. The use and effects of electronic health tools for patient self-monitoring and reporting of outcomes following medication use: systematic review. *J Med Internet Res* 2018 Dec 18;20(12):e294 [FREE Full text] [doi: [10.2196/jmir.9284](https://doi.org/10.2196/jmir.9284)] [Medline: [30563822](https://pubmed.ncbi.nlm.nih.gov/30563822/)]
12. Chaimowitz GA, Mamak M, Moulden HM, Furimsky I, Olagunju AT. Implementation of risk assessment tools in psychiatric services. *J Healthc Risk Manag* 2020 Jul;40(1):33-43. [doi: [10.1002/jhrm.21405](https://doi.org/10.1002/jhrm.21405)] [Medline: [32301249](https://pubmed.ncbi.nlm.nih.gov/32301249/)]
13. Bowen SJ, Graham ID. From knowledge translation to engaged scholarship: promoting research relevance and utilization. *Arch Phys Med Rehabil* 2013 Jan;94(1 Suppl):S3-S8. [doi: [10.1016/j.apmr.2012.04.037](https://doi.org/10.1016/j.apmr.2012.04.037)] [Medline: [23141502](https://pubmed.ncbi.nlm.nih.gov/23141502/)]
14. Wilkinson H, Gallagher M, Smith M. A collaborative approach to defining the usefulness of impact: lessons from a knowledge exchange project involving academics and social work practitioners. *Evidence Policy* 2012 Aug 1;8(3):311-327. [doi: [10.1332/174426412X654040](https://doi.org/10.1332/174426412X654040)]
15. Majid S, Reeves S, Figueredo G, Brown S, Lang A, Moore M, et al. The extent of user involvement in the design of self-tracking technology for bipolar disorder: literature review. *JMIR Ment Health* 2021 Dec 20;8(12):e27991 [FREE Full text] [doi: [10.2196/27991](https://doi.org/10.2196/27991)] [Medline: [34931992](https://pubmed.ncbi.nlm.nih.gov/34931992/)]
16. Gagliardi AR, Berta W, Kothari A, Boyko J, Urquhart R. Integrated knowledge translation (IKT) in health care: a scoping review. *Implement Sci* 2016 Mar 17;11:38 [FREE Full text] [doi: [10.1186/s13012-016-0399-1](https://doi.org/10.1186/s13012-016-0399-1)] [Medline: [26988000](https://pubmed.ncbi.nlm.nih.gov/26988000/)]
17. Wensing M, Grol R. Knowledge translation in health: how implementation science could contribute more. *BMC Med* 2019 May 07;17(1):88 [FREE Full text] [doi: [10.1186/s12916-019-1322-9](https://doi.org/10.1186/s12916-019-1322-9)] [Medline: [31064388](https://pubmed.ncbi.nlm.nih.gov/31064388/)]
18. Szram J, Cullinan P. Medical surveillance for prevention of occupational asthma. *Curr Opin Allergy Clin Immunol* 2013 Apr;13(2):138-144. [doi: [10.1097/ACI.0b013e32835e1212](https://doi.org/10.1097/ACI.0b013e32835e1212)] [Medline: [23339935](https://pubmed.ncbi.nlm.nih.gov/23339935/)]
19. Reynolds LE, Wolfe AL, Clark KA, Blackley DJ, Halldin CN, Laney AS, et al. Strengthening the coal workers' health surveillance program. *J Occup Environ Med* 2017 Apr;59(4):e71 [FREE Full text] [doi: [10.1097/JOM.0000000000000993](https://doi.org/10.1097/JOM.0000000000000993)] [Medline: [28628061](https://pubmed.ncbi.nlm.nih.gov/28628061/)]

20. Nafees AA, Fatmi Z. Available interventions for prevention of cotton dust-associated lung diseases among textile workers. *J Coll Physicians Surg Pak* 2016 Aug;26(8):685-691. [Medline: [27539764](#)]
21. Malo JL, L'Archevêque J, Trudeau C, d'Aquino C, Cartier A. Should we monitor peak expiratory flow rates or record symptoms with a simple diary in the management of asthma? *J Allergy Clin Immunol* 1993 Mar;91(3):702-709. [doi: [10.1016/0091-6749\(93\)90189-m](#)] [Medline: [8454792](#)]
22. Archibald MM, Hartling L, Ali S, Caine V, Scott SD. Developing "My Asthma Diary": a process exemplar of a patient-driven arts-based knowledge translation tool. *BMC Pediatr* 2018 Jun 05;18(1):186 [FREE Full text] [doi: [10.1186/s12887-018-1155-2](#)] [Medline: [29871611](#)]
23. Carpenter DM, Geryk LL, Sage A, Arrindell C, Sleath BL. Exploring the theoretical pathways through which asthma app features can promote adolescent self-management. *Transl Behav Med* 2016 Dec;6(4):509-518 [FREE Full text] [doi: [10.1007/s13142-016-0402-z](#)] [Medline: [27118115](#)]
24. Morita PP, Yeung MS, Ferrone M, Taite AK, Madeley C, Stevens Lavigne A, et al. A patient-centered mobile health system that supports asthma self-management (breathe): design, development, and utilization. *JMIR Mhealth Uhealth* 2019 Jan 28;7(1):e10956 [FREE Full text] [doi: [10.2196/10956](#)] [Medline: [30688654](#)]
25. To T, Cicutto L, Degani N, McLimont S, Beyene J. Can a community evidence-based asthma care program improve clinical outcomes?: a longitudinal study. *Med Care* 2008 Dec;46(12):1257-1266. [doi: [10.1097/MLR.0b013e31817d6990](#)] [Medline: [19300316](#)]
26. Olajos-Clow J, Szpiro K, Julien B, Minard J, Loughheed MD. Emergency department adult asthma care pathway: healthcare providers' perceived utility and barriers to implementation. *Adv Emerg Nurs J* 2009;31(1):44-53. [doi: [10.1097/TME.0b013e3181946f76](#)] [Medline: [20118853](#)]
27. Killorn KR, Dostaler SM, Groome PA, Loughheed MD. The use of a work-related asthma screening questionnaire in a primary care asthma program: an intervention trial. *J Asthma* 2015 May;52(4):398-406. [doi: [10.3109/02770903.2014.971966](#)] [Medline: [25290815](#)]
28. Killorn KR, Dostaler SM, Olajos-Clow J, Turcotte SE, Minard JP, Holness DL, et al. The development and test re-test reliability of a work-related asthma screening questionnaire. *J Asthma* 2015 Apr;52(3):279-288. [doi: [10.3109/02770903.2014.956892](#)] [Medline: [25180965](#)]
29. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](#)]
30. Berger R. Now I see it, now I don't: researcher's position and reflexivity in qualitative research. *Qual Res* 2013 Jan 03;15(2):219-234. [doi: [10.1177/1468794112468475](#)]
31. Holness DL, Tabassum S, Tarlo SM, Liss GM, Silverman F, Manno M. Practice patterns of pulmonologists and family physicians for occupational asthma. *Chest* 2007 Nov;132(5):1526-1531. [doi: [10.1378/chest.06-2224](#)] [Medline: [17890481](#)]
32. Ibrahim F, Sandström P, Björnsson B, Larsson AL, Drott J. 'I want to know why and need to be involved in my own care...': a qualitative interview study with liver, bile duct or pancreatic cancer patients about their experiences with involvement in care. *Support Care Cancer* 2019 Jul;27(7):2561-2567 [FREE Full text] [doi: [10.1007/s00520-018-4548-8](#)] [Medline: [30430301](#)]
33. Tamirisa NP, Goodwin JS, Kandalam A, Linder SK, Weller S, Turrubiate S, et al. Patient and physician views of shared decision making in cancer. *Health Expect* 2017 Dec;20(6):1248-1253 [FREE Full text] [doi: [10.1111/hex.12564](#)] [Medline: [28464430](#)]
34. Boulet LP, FitzGerald JM, Levy ML, Cruz AA, Pedersen S, Haahtela T, et al. A guide to the translation of the Global Initiative for Asthma (GINA) strategy into improved care. *Eur Respir J* 2012 May;39(5):1220-1229 [FREE Full text] [doi: [10.1183/09031936.00184511](#)] [Medline: [22282546](#)]
35. Lippel K. Workers describe the effect of the workers' compensation process on their health: a Québec study. *Int J Law Psychiatry* 2007;30(4-5):427-443. [doi: [10.1016/j.ijlp.2007.06.013](#)] [Medline: [17692917](#)]
36. Qin J, Kurowski A, Gore R, Punnett L. The impact of workplace factors on filing of workers' compensation claims among nursing home workers. *BMC Musculoskelet Disord* 2014 Jan 29;15:29 [FREE Full text] [doi: [10.1186/1471-2474-15-29](#)] [Medline: [24476529](#)]
37. Walters GI, Soundy A, Robertson AS, Burge PS, Ayres JG. Understanding health beliefs and behaviour in workers with suspected occupational asthma. *Respir Med* 2015 Mar;109(3):379-388 [FREE Full text] [doi: [10.1016/j.rmed.2015.01.003](#)] [Medline: [25657173](#)]
38. Bradshaw LM, Barber CM, Davies J, Curran AD, Fishwick D. Work-related asthma symptoms and attitudes to the workplace. *Occup Med (Lond)* 2007 Jan;57(1):30-35. [doi: [10.1093/occmed/kql097](#)] [Medline: [17200128](#)]
39. Garvey CM, Jones R. Is there a place for theoretical frameworks in qualitative research? *Int J Qual Methods* 2021 Feb 27;20:160940692098795. [doi: [10.1177/1609406920987959](#)]
40. Morse JM. The power of induction. *Qual Health Res* 1992 Feb;2(1):3-6. [doi: [10.1177/104973239200200101](#)]

## Abbreviations

ARU: Asthma Research Unit

**EMR:** electronic medical record

**KTA:** knowledge-to-action

**PAC:** project advisory committee

**WRA:** work-related asthma

**WRASQ(L):** Work-related Asthma Screening Questionnaire–Long version

*Edited by A Mavragani; submitted 23.02.22; peer-reviewed by J Lowery, D Madhusudhan; comments to author 27.06.22; revised version received 03.08.22; accepted 10.08.22; published 15.09.22.*

*Please cite as:*

*MacKinnon M, Moloney M, Bullock E, Morra A, To T, Lemiere C, Lougheed MD*

*Implementation of a Work-Related Asthma Screening Questionnaire in Clinical Settings: Multimethods Study*

*JMIR Form Res 2022;6(9):e37503*

*URL: <https://formative.jmir.org/2022/9/e37503>*

*doi: [10.2196/37503](https://doi.org/10.2196/37503)*

*PMID: [35964327](https://pubmed.ncbi.nlm.nih.gov/35964327/)*

©Madison MacKinnon, Max Moloney, Emma Bullock, Alison Morra, Teresa To, Catherine Lemiere, M Diane Lougheed. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 15.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Digital Platform to Continuously Monitor Patients Using a Smartwatch: Preliminary Report

Kaio Jia Bin<sup>1</sup>, MD, MSc; Lucas Ramos De Pretto<sup>1</sup>, PhD; Fabio Beltrame Sanchez<sup>1</sup>, MSc; Linamara Rizzo Battistella<sup>1</sup>, MD, PhD

Instituto de Medicina Física e Reabilitação do Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil

**Corresponding Author:**

Kaio Jia Bin, MD, MSc  
Instituto de Medicina Física e Reabilitação do Hospital das Clínicas  
Faculdade de Medicina  
Universidade de São Paulo  
Rua Domingo Soto 100  
Jardim Vila Mariana  
São Paulo  
Brazil  
Phone: 55 1126616208  
Email: [kaiobin@gmail.com](mailto:kaiobin@gmail.com)

## Abstract

**Background:** Monitoring vital signs such as oximetry, blood pressure, and heart rate is important to follow the evolution of patients. Smartwatches are a revolution in medicine allowing the collection of such data in a continuous and organic way. However, it is still a challenge to make this information available to health care professionals to make decisions during clinical follow-up.

**Objective:** This study aims to build a digital solution that displays vital sign data from smartwatches, collected remotely, continuously, reliably, and from multiple users, with trigger warnings when abnormal results are identified.

**Methods:** This is a single-center prospective study following the guidelines “Evaluating digital health products” from the UK Health Security Agency. A digital platform with 3 different applications was created to capture and display data from the mobile phones of volunteers with smartwatches. We selected 80 volunteers who were followed for 24 weeks each, and the synchronization interval between the smartwatch and digital solution was recorded for each vital sign collected.

**Results:** In 14 weeks of project progress, we managed to recruit 80 volunteers, with 68 already registered in the digital solution. More than 2.8 million records have already been collected, without system downtime. Less than 5% of continuous heart rate measurements (bpm) were synchronized within 2 hours. However, approximately 70% were synchronized in less than 24 hours, and 90% were synchronized in less than 119 hours.

**Conclusions:** The digital solution is working properly in its role of displaying data collected from smartwatches. Vital sign values are being monitored by the research team as part of the monitoring of volunteers. Although the digital solution proved unsuitable for monitoring urgent events, it is more than suitable for use in outpatient clinical use. This digital solution, which is based on cloud technology, can be applied in the future for telemonitoring in regions lacking health care professionals. Accuracy and reliability studies still need to be performed at the end of the 24-week follow-up.

(*JMIR Form Res* 2022;6(9):e40468) doi:[10.2196/40468](https://doi.org/10.2196/40468)

**KEYWORDS**

smartwatch; digital health; telemedicine; wearable; telemonitoring; mobile health; digital platform; clinical intervention; sensitive data; clinical trial

## Introduction

**Background**

Since the World Health Organization declared the novel coronavirus a pandemic on March 11, 2020 [1], with more than

470 million cases of infection and more than 6 million deaths confirmed [2], digital transformation of health care worldwide has accelerated [3-5].

In this scenario where the most frequent comorbidities are hypertension (55%), coronary artery disease and stroke (32%),

and diabetes (31%) [6], monitoring vital signs such as oximetry, blood pressure, and heart rate can be of paramount importance to monitor the evolution of patients infected by COVID-19.

Thus, wearable devices, such as smartwatches, are key actors in revolutionizing medicine through mobile health (mHealth) and eHealth, allowing continuous and longitudinal health monitoring outside of health care facilities [7].

Due to the ease of use of smartwatches—initially aimed at consumers concerned about their own health—several studies have shown interest in their application for remote monitoring [8] and as a tool for telemonitoring and early detection of respiratory symptoms [9,10], heart disease [11-14], and remote physical therapy [15].

Smartwatches can record clinical data in a way that feels organic and unobtrusive to the user, enabling the construction of a database that will facilitate, with the aid of artificial intelligence, the recognition of biomarkers capable of expanding the mechanisms of prediction, prevention, and health event intervention.

There are reports in the literature discussing smartwatch data collection [16], with the most recent already suggesting the possibility of early detection of COVID-19 [9,10] and atrial fibrillation [12-14]. However, for wearable devices to be really usable in a clinical setting, a digital platform that is easily accessible to health professionals is necessary.

### Motivation

Aiming to enable the future of telemonitoring, early detection, and remote therapies, there is a need for a digital web platform capable of collecting data from smartwatches continuously and effectively.

### Aim

This study aimed to build a digital solution that displays vital sign data from smartwatches, collected remotely, continuously,

reliably, without the need on manual input, and from multiple users, triggering warnings when abnormal results are identified.

## Methods

The digital solution was developed to support an entire clinical study involving remote monitoring of patients. Its role is to guarantee anonymity, completeness, and reliability of the data collected, as well as to consolidate different sources of information input.

### Ethical Considerations

This project was submitted to the Ethics and Research Committee of the Hospital das Clinicas da Faculdade de Medicina of University of São Paulo (HCFMUSP; CAAE: 51711921.3.0000.0068 and Opinion number: 4,975,512).

### Study Type

This is a single-center prospective study following the guidelines “Evaluating digital health products” from the UK Health Security Agency [17], with local adaptations for the Brazilian population and project context, still in the design and test phases of a digital product.

A descriptive study, “Analysis of routinely collected data” [18], was conducted using data collected from the digital solution to a website platform by the team of executing researchers.

### Data Transfer

There are several ways to collect data from a smartwatch to a digital platform, as described by de Arriba-Pérez et al [19]. To ensure the validity and protection of the volunteer’s sensitive data [20], we developed an application to be used in parallel with the manufacturer’s app.

In this paper, only data used for clinical follow-up of volunteers were selected, namely blood pressure, oxygen saturation, heart rate, and sleep quality information (Table 1).

**Table 1.** Information collected from the smartwatch for this study.

Data	Collection	Present at digital solution
Steps	Automatically	No
Flights of stairs	Automatically	No
Exercise time	Automatically	No
Sleep quality	Automatically	Yes
Heart rate	Automatically	Yes
Oxygen saturation during sleep	Automatically	No
Oxygen saturation	User action	Yes
Blood pressure	User action	Yes
Weight	User action	No
Height	User action	No
Quantity and type of liquid ingested	User action	No
Quantity and type of food ingested	User action	No

**Data Flow**

Vital sign data are captured by the smartwatch during its use by the volunteer. The smartwatch then synchronizes with the volunteer’s smartphone, and the data are transferred to the manufacturers’ health apps.

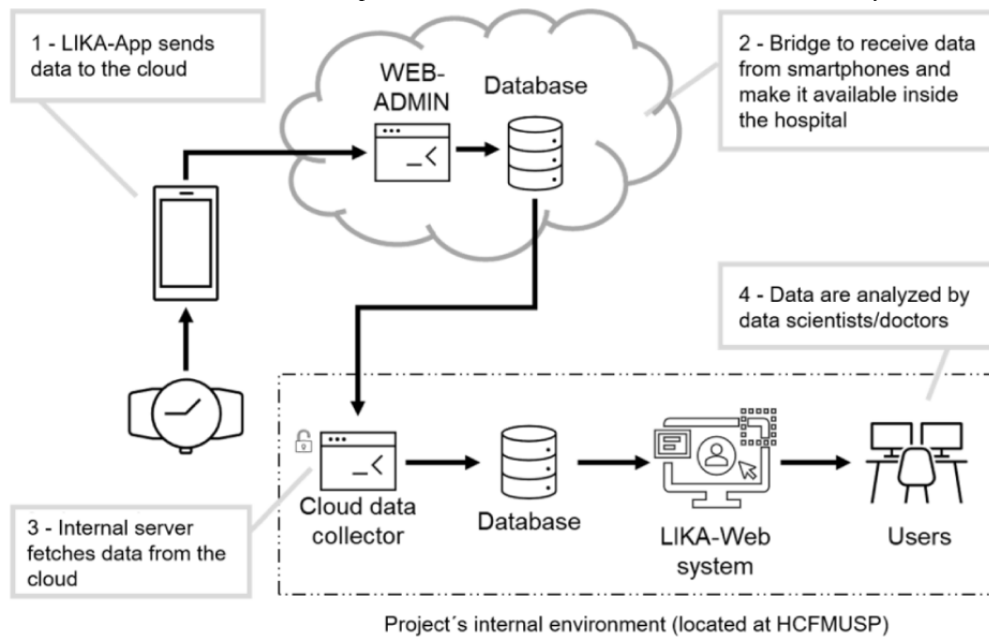
A custom application, called LIKA-App, reads the data from those health apps and transfers the data to a cloud server. A cloud data collector installed on the internal server of HCFMUSP downloads the data from the cloud server.

A second, complementary digital solution, LIKA-Web, makes the acquired data available for analysis and allows alerts to be set up through a user-friendly interface.

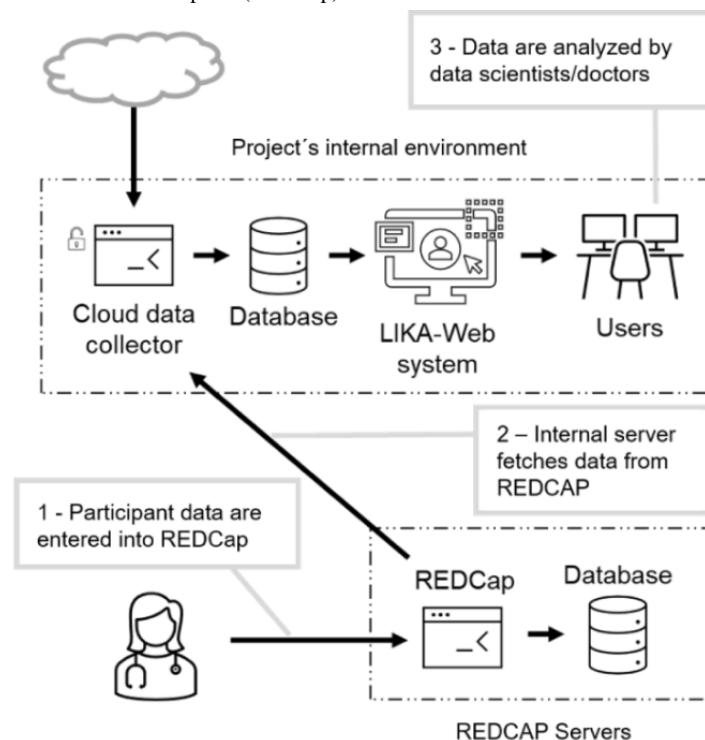
The data flow is shown in Figure 1.

In addition to those collected by the smartwatch, data obtained by gold standard devices will also be persisted in the project’s internal server database. These data will be entered manually in the Research Electronic Data Capture (REDCap) [21] system and later imported and incorporated into the study’s database. A diagram to visualize the integration between REDCap and the LIKA-Web solution can be found in Figure 2.

**Figure 1.** Data flow from smartwatches. HCFMUSP: Hospital das Clinicas da Faculdade de Medicina of University of São Paulo.



**Figure 2.** Data flow from Research Electronic Data Capture (REDCap).



## Digital Solutions

### *LIKA-App*

This app is a digital solution installed on volunteers' smartphones that captures specific health information from the wearable device and synchronizes it with the study's data cloud (WEB-ADMIN server). This application runs on Android smartphones and was developed and improved by the manufacturer to accommodate the requirements of this study.

### *WEB-ADMIN*

This is the module responsible for receiving data obtained from the LIKA-App and storing it in the study's data cloud for later use by the LIKA-Web. It also tracks the date and time of the latest synchronization of each participating volunteer's smartwatch. This module provides the REST APIs [22] consumed by LIKA-App and a structured database (PostgreSQL).

### *LIKA-Web*

LIKA-Web is the module responsible for supporting the remote monitoring project. It is a digital solution capable of retrieving information from different data sources for each volunteer. A user-friendly interface allows the team of researchers and clinicians to follow the day-to-day activities of the clinical study, as well as receiving alerts about vital signs and synchronization information.

To allow fast development and good reliability, LIKA-Web was built using Java Spring Boot framework (back-end) and Bootstrap/HTML/JQuery (front-end), with a MySQL database.

## Data Sync

The synchronization of data between the manufacturer's health app and the LIKA-App takes place either under a deliberate request by the user, at any time, or automatically, once a day, but only while the app is active on the user's phone. This synchronization is dependent on internet connectivity.

Once the LIKA-App retrieves information from the manufacturer's application, these data are available to be synchronized with LIKA-Web, which takes place automatically via WEB-ADMIN every hour or under deliberate request by the Web Administrator at any time.

## Functionality

The platform offers the functionality described in the following paragraphs.

User registration grants users access to the LIKA-Web system.

During volunteer registration, for privacy protection, the volunteer's name is not registered in our internal system (LIKA-Web). Instead, a unique user identifier is created and registered in the LIKA-App mobile app.

Volunteer risk alerts are triggered to warn the health care team based on the data received from the volunteers.

Data sync alerts identify volunteers who have not synced their smartwatch data in the last 72 hours.

The latest data log is an interface that logs the latest data obtained from each volunteer, along with the date and hour.

The data display interface displays the vital sign data obtained from the smartwatches.

## Wearable and Gold Standard Device

Each volunteer was provided with a smartwatch (Samsung Galaxy 4). The gold standard devices used in this study are a noninvasive blood pressure monitor, (G-TECH model GP400 [ANVISA registration n° 80275319016] with 2 AAA batteries) and a pulse oximeter (AFK YK009 [ANVISA registration 81995169005]) for noncontinuous monitoring.

A smartphone (Samsung A52) was provided to volunteers who did not have a Samsung mobile phone.

## Data Settings

Among the HCFMUSP collaborators, 80 volunteers were selected to use the smartwatch over 24 weeks, with daily visits by research monitors.

The following data were collected from the smartwatch by the LIKA-App: heart rate, oxygen saturation, blood pressure, sleep count, and sleep intensity.

Heart rate is configured to be collected continuously in 1-minute intervals. The smartwatch sends the value in bpm to the LIKA digital solution.

Oxygen saturation can be collected manually by the user or continuously during sleep.

Blood pressure is collected upon user request, and the smartwatch needs to be calibrated with a gold standard blood pressure device once every 28 days.

Sleep count is a feature that marks the beginning and end of the volunteer's sleep period. Sleep intensity data record the beginning and end of each sleep phase [23]. Both types of sleep data are automatically collected by the smartwatch without the need for user intervention.

For each of these values, the digital platform also collects the date and time of their record by the smartwatch, as well as the date and time when the digital platform synchronized and received the data.

Furthermore, every day, for 24 weeks, a research monitor meets with the volunteers and simultaneously collects oxygen saturation and blood pressure data from both the volunteer's smartwatch and gold standard devices. Those values, along with date and beginning and end times (as displayed by the smartwatch), are registered on a paper form and later transcribed to REDCap, in order to analyze the reliability and agreement between the gold standard and smartwatch data.

For the purpose of this study, the automatic synchronization interval between LIKA-Web and LIKA-App is set to every 1 hour.

## Digital Platform Reliability Analysis

To determine the reliability of the data captured by the LIKA platform (as seen in [Figure 1](#)) at the end of the study, the data

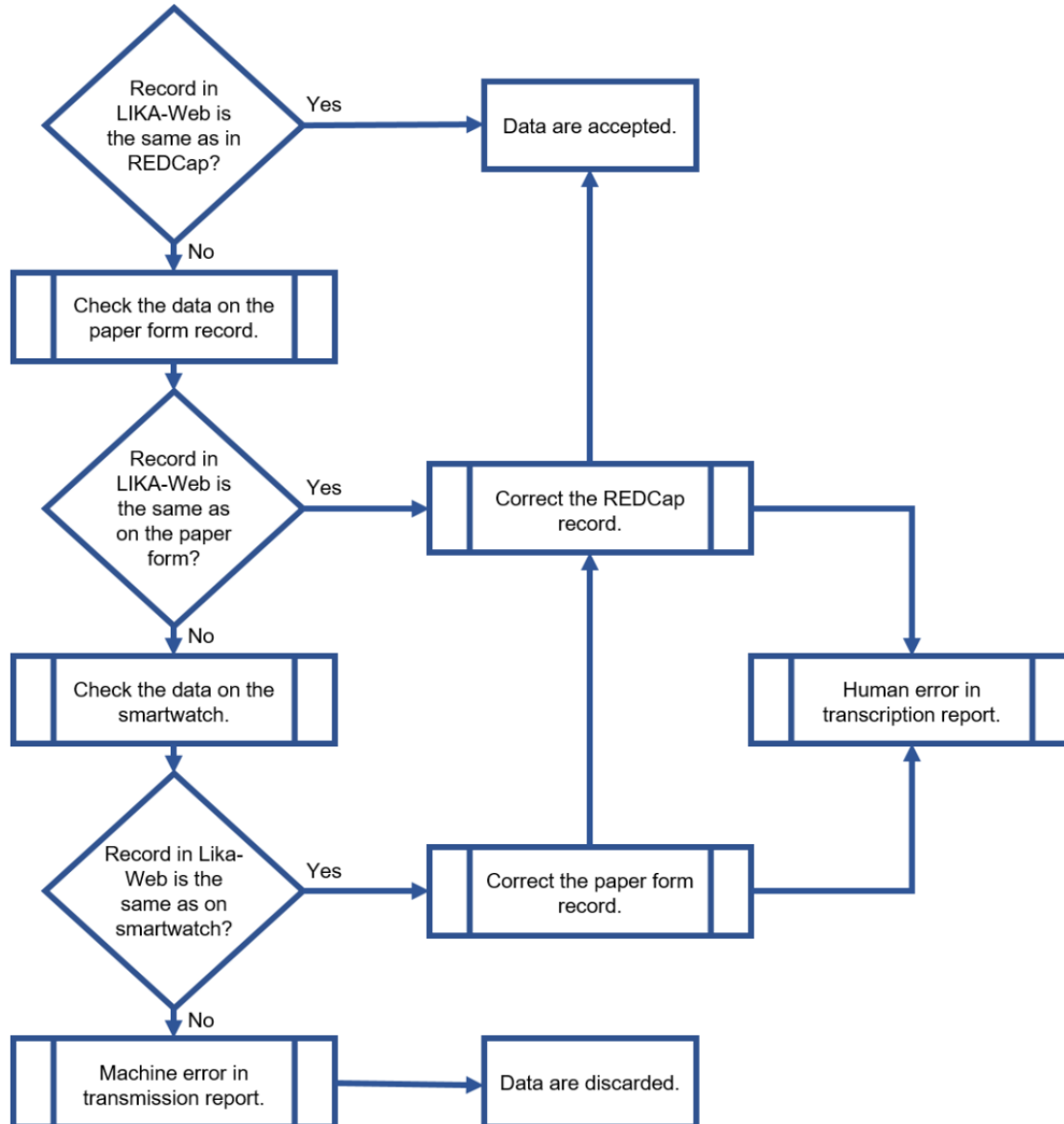


collected from the smartwatch and manually recorded in REDCap will be compared with those directly captured by WEB-ADMIN.

An active search will be carried out by combining data manually annotated in REDCap, by day, time, and volunteer, within the internal server database.

The comparison flow is in [Figure 3](#).

**Figure 3.** Comparison of the flow of data collected by the digital platform with those recorded in Research Electronic Data Capture (REDCap), on the paper form, and by smartwatch.



### Alert System Measurement

The following triggers were defined for the blood pressure alerts, based on the manufacturer's recommendations [24]: systolic blood pressure  $\leq 70$  mm Hg or  $> 180$  mm Hg and diastolic blood pressure  $\leq 40$  mm Hg or  $> 120$  mm Hg.

Every time one of these events is triggered by the digital platform, the value, date, hour, and volunteer's identifier are displayed on the dashboard.

With the intention of monitoring the use of the smartwatch by the volunteers, alerts regarding oxygen saturation and heartbeat per minute were also implemented on the digital platform.

Considering the normal oxygen levels in a pulse oximeter usually range from 95% to 100% and hypoxemia is an oxygen saturation of less than 90% [25], the digital platform was set up to alert when the value reaches below 88%.

Similarly, considering that the diagnosis of sinus bradycardia requires an electrocardiogram showing a normal sinus rhythm at a rate lower than 60 bpm [26], the digital platform was set up to alert when the heart rate falls below 40 bpm.

### Results

Running without interruption since February 25, 2022, a total of 68 volunteers were already being monitored by the

LIKA-Web platform as of May 28, 2022 (Table 2), with 3 dropouts in the first week of the study.

Among the dropouts, a lack of time to attend the daily face-to-face data collection visits was the leading reason for abandoning the study. Data from these 3 dropout volunteers (465 records) were excluded from the analysis.

A total of 2,772,766 records were captured: 2,645,286 for continuous heart rate, 7423 for oxygen saturation, 4742 for blood pressure, 3599 for sleep count (marks the beginning and end of sleep), and 111,716 for sleep intensity (marks the beginning and end of each phase of sleep).

There were 9 users with continuous heart bpm data synchronization issues, 1 with oxygen saturation data issues, and 10 with blood pressure data synchronization issues (Table 3).

We noticed that many heart rate synchronization issues were due to the smartwatch battery management configuration: The smartwatches provided to the volunteers were configured in the continuous acquisition mode; however, as soon as the smartwatch entered battery saving mode, the configuration reverted back to only 1 measurement every 10 minutes. To

address this problem, the research team advised volunteers not to use the battery saving mode on their devices.

Regarding the oxygen saturation issues, there was a synchronization problem with a single, newly recruited volunteer, and the research team was working to identify the reason.

Finally, for the blood pressure synchronization issues, we verified that the cause was a sharing permission between Samsung Monitor and Samsung Health App. The devices were configured with all permissions when the smartwatches were handed out to the volunteers; however, for unconfirmed reasons (likely system updates), the permission configurations were lost. To solve this problem, the research team reconfigured the data sharing permission as soon as the problem was identified by our data sync alert system.

Based on the collected data, descriptive statistics were performed across all the records obtained (Table 4). The overall results for heart rate and oxygen saturation show a distribution (between the 25th and 75th percentiles) within the expected normative values, indicating that there were no persistent abnormal readings from our volunteers.

**Table 2.** Total of users by weeks of the project.

Week number	New users, n	Number of users who dropped out, n	Total users, n
1	2	0	2
2	7	0	9
3	8	0	17
4	5	0	22
5	9	0	31
6	6	0	37
7	4	1	40
8	9	1	48
9	2	1	49
10	7	0	56
11	3	0	59
12	2	0	61
13	6	0	67
14	1	0	68

**Table 3.** Total data records by type and per user (N=2,772,766).

Data type	Records, n	Total users, n	Users with no synchronization, n
Continuous heart rate (bpm)	2,645,286	68	9
Oxygen saturation	7423	68	1
Blood pressure	4742	68	10
Sleep	3599	68	0
Sleep intensity	111,716	68	0

**Table 4.** Descriptive statistics of all the collected vital signs.

Data type	Results				
	Mean (SD)	95% CI	25th percentile	50th percentile	75th percentile
Heart rate (bpm)	78.29 (13.63)	78.27-78.30	69.00	77.00	86.00
Oxygen saturation (%)	94.37 (3.47)	94.29-94.45	93.00	95.00	97.00
<b>Blood pressure (mm Hg)</b>					
Systolic	126.78 (15.06)	126.35-127.21	116.00	127.00	136.00
Diastolic	82.45 (12.16)	82.10-82.79	73.00	83.00	92.00

Similarly, the blood pressure results present a distribution mostly inside the range expected for resting blood pressure. Individual averages of heart rate for each patient also present a normative distribution, with lower (Q1), median (Q2), and upper (Q3) quartile values of 74.57 bpm, 78.24 bpm, and 82.66 bpm, respectively. The distribution of the individual means for oxygen saturation is also compatible with the normative range but shifted to slightly lower values: 93.67%, 94.64%, and 95.24% (Q1, Q2, and Q3, respectively).

The same observations were made for individual averages for systolic (Q1: 116.69 mm Hg; Q2: 120.50 mm Hg; Q3: 129.06 mm Hg) and diastolic (Q1: 76.16 mm Hg; Q2: 82.16 mm Hg; Q3: 89.63 mm Hg) blood pressures.

Data regarding sleep intensity were also analyzed. The smartwatch records sleep intensity at 4 different levels: still awake, light sleep, deep sleep, and rapid eye movement (REM).

The total sleep time recorded was verified, and the percentage of this total spent at each sleep intensity was calculated. The results are shown in [Table 5](#), where it can be seen that more than one-half (590,433/1,046,195, 56.43%) of the recorded sleep time was spent in light sleep, and only about 20% (213,716/1,046,195, 20.43%) was spent in REM sleep. Yet, despite this, we found that all volunteers managed to achieve REM sleep at least once.

Still, regardless of the distributions in [Table 3](#), a few vital sign records fell outside the range of normative values, and, to monitor those, the alert system within our framework was used. In all, 421 alerts were generated ([Table 6](#)).

**Table 5.** Time spent in each sleep stage across all recorded minutes (1,046,195 minutes).

Sleep phase	Overall recorded minutes (% of total)
Awake	122,898 (11.75)
Light sleep	590,433 (56.43)
Deep sleep	119,148 (11.39)
REM <sup>a</sup> sleep	213,716 (20.43)

<sup>a</sup>REM: rapid eye movement.

To assess whether the platform can be used by the clinical team as a digital health care solution, it is very important that the data be available in the digital platform as soon as possible after being recorded by the smartwatch.

The time interval between the moment of recording the continuous beats per minute (cBPM) and the moment the value is available on the digital platform depends on the synchronization between the manufacturer's health app and LIKA-Web, via the LIKA-App.

In total, 34 Samsung A52 smartphones were provided to volunteers who did not already have a Samsung mobile phone. In addition, 34 volunteers provided their own Samsung smartphone (multiple models) for use in this study. The proportions of cases with synchronization issues were similar, regardless of whether the volunteer was using their own smartphone ([Table 7](#)).

As shown in [Table 8](#), less than 5% of records synced within the first 2 hours of interval time, either in the group with their own smartphone (55,955/1,349,706, 4.15%) or in the group with a smartphone provided by the study (64,943/1,295,580, 5.01%). On the other hand, most of the data were synchronized within 24 hours in both groups, and only less than 10% of data were synced after 7 days.

So far, no downtime events have been recorded on the LIKA-Web platform, and this project is scheduled to end on October 23, 2022.

**Table 6.** Type and number of alerts and number of volunteers who triggered each alert.

Type of alert	Number of alerts (n=421)	Number of volunteers (n=45)
Maximum diastolic blood pressure ( $\geq 180$ mm Hg)	1	1
Minimum diastolic blood pressure ( $\leq 70$ mm Hg)	0	0
Maximum systolic blood pressure ( $\geq 120$ mm Hg)	3	1
Minimum systolic blood pressure ( $\leq 40$ mm Hg)	0	0
Low oxygen saturation ( $< 88\%$ )	414	43
Low heart rate ( $< 40$ bpm)	3	2

**Table 7.** Number of volunteers with synchronization (sync) errors.

Volunteers	Sync OK (n=59), n	Sync error (n=9), n	cBPM <sup>a</sup> records (n=2,645,286), n
With own smartphone	30	4	1,349,706
With study's smartphone	29	5	1,295,580

<sup>a</sup>cBPM: continuous beats per minute.

**Table 8.** Synchronization time according to the smartphone used.

Sync time	Study's smartphone (n=1,295,580)		Volunteer's smartphone (n=1,349,706)	
	Number synced, n (%)	Accumulated %	Number synced, n (%)	Accumulated %
<1 hour	5051 (0.39)	0.39	4400 (0.33)	0.33
1-2 hours	59,892 (4.62)	5.01	51,555 (3.82)	4.15
3-6 hours	193,524 (14.94)	19.95	166,518 (12.34)	16.48
6-12 hours	278,549 (21.50)	41.45	266,397 (19.74)	36.22
13-24 hours	349,405 (26.97)	68.42	452,547 (33.53)	69.75
1-2 days	143,495 (11.08)	79.49	196,683 (14.57)	84.32
3-7 days	154,646 (11.94)	91.43	131,064 (9.71)	94.03
7-14 days	78,115 (6.03)	97.46	55,495 (4.11)	98.14
>14 days	32,903 (2.54)	100	25,047 (1.86)	100

## Discussion

### Principal Findings

The LIKA-Web platform has been stable and is ready to fulfill its role of receiving data from mobile devices.

In the 14 weeks of follow-up as of the writing of this manuscript, we had not had any interruption of its service, and it is being actively used to monitor the health of the volunteers as well as to verify if they performed the daily synchronization with the platform.

In the first week of recruitment, the equipment was delivered to the volunteers on the day of their consultation, which incurred a lot of difficulty with the setup time. It took more than 90 minutes just to charge the smartwatch, pair it with the mobile phone, and install the apps.

From the second week onwards, equipment delivery was scheduled before the consultations, with a pairing and installation guide, which greatly reduced the time of those consultations.

More than 2.8 million records were received by the platform, and the expectation is to reach more than 20 million data records by the end of the project on October 23, 2022.

Reliability studies on data transmission will be carried out as reported in the Methods section, and other studies will also make use of this large mass of data.

As the volunteers were also grouped by disease (COVID-19) status and demographic characteristics, we expect to enable analyses regarding the evolution of COVID-19 and accuracy of the smartwatch with respect to the gold standard equipment, among other possibilities that we are following on a daily basis.

During the follow-up medical appointments during our study, through the digital platform, it was possible to identify a pattern of asymptomatic bradycardia in one of the volunteers, for whom the only symptom was tiredness.

So far, 2 volunteers without previous positive results for COVID-19 were infected by the virus during the study, staying in home isolation and accompanied by the digital platform.

All these data, added to the participants' individual characteristics and clinical follow-up, will be evaluated and

analyzed at the end of the study, thanks to the digital platform that makes the collected data available for continuous monitoring, at a distance, without the need for the intervention of a health professional.

Compared with similar recent studies by Quer et al [9] and Mishra et al [10] that approached the detection of COVID-19 through the use of heart rate, steps, and sleep data, the digital solution in this study also provides researchers and clinicians data about oxygenation during sleep [27], in an organic and continuous way, in addition to data on blood pressure and oxygenation, upon volunteer request.

Furthermore, as very well detailed by Vijayan et al [28], smartwatches have very broad applicability in the health care field. Since the LIKA digital solution was developed together with the manufacturer, it is ready to receive and provide, for clinical use, data from other features not included in this version of the study, as is the case for electrocardiogram for atrial fibrillation monitoring reported in the studies by Nasarre et al [11], Bumgarner et al [12], and Perez et al [13].

For the purpose of clinical use, with no intervention to encourage volunteers to sync the data and automatic synchronization between the digital platform (LIKA-Web) and digital solution on the mobile phone (LIKA-App) set to every 1 hour, the total time delay in synchronization was under 1 hour for less than 0.5% of the records and under 2 hours for only 5% of the records (Table 8).

These results indicate that this solution is not suitable for monitoring critical events, which makes the alert system unsuitable for urgent clinical use.

However, almost 70% of the data are synchronized within 24 hours (Table 8), and less than 10% of the data are synchronized in more than 7 days (Table 8), indicating that the digital solution is more than adequate to be used as a nonurgent health care tool.

Given the LIKA platform was developed with scalability in mind, it is suitable to be used in multicenter, large-scale, and

long-term follow-up studies, and it can provide researchers and clinicians with a powerful tool to monitor their patients.

In addition, as the digital solution LIKA is entirely in Brazilian Portuguese and based on cloud technology, it is ready to be accessed in any region of a continental country like Brazil, a region that still lacks studies in this area. Regions in need of medical assistance and located in hard-to-reach areas, such as Amazonas, will be able to take advantage of this technology, as suggested in recent studies carried out by HCFMUSP [29]. Those regions may greatly benefit from a system that enables patient follow-up remotely, such as the one proposed here.

### Limitations

This work is a preliminary report of an ongoing study.

Data are still insufficient for analysis of reliability and accuracy of smartwatches. However, the platform has proven itself ready for data collection on a large scale, and the vital sign values have been used by the medical team in return visits to guide health assessment.

Without in-depth analysis, some records with human error in the data transcription to paper form or to REDCap were identified and confirmed; that is, the data on the digital platform are identical to those recorded on the smartwatches but different from those inputted on the paper form and in REDCap.

Furthermore, this paper reports only part of the project “Predictive monitoring in person-centered care using smartwatches.” Methodologies about the division of the groups of volunteers, clinical follow-up setting, clinical scale used to evaluate COVID-19 patients, and user satisfaction surveys are not part of this preliminary report, as they are not relevant to the operating results of the digital platform,

### Cost

In all, US \$39,053.98 (BRL to US \$ conversion of 5.80 obtained on December 31, 2021) were spent on smartwatches, gold standard devices, and mobile phones in this project (Table 9).

**Table 9.** Total investment on devices and equipment.

Device	Quantity, n	Investment (US \$)
Samsung Galaxy 4 smartwatch	84	19,828.73
Noninvasive blood pressure monitor	92	1866.57
Pulse oximeter for noncontinuous monitoring	92	1130.24
Samsung Galaxy A52 smartphone	45	16,228.44

### Conclusions

The LIKA-Web platform is working correctly in its role of presenting data collected from smartwatches worn by research volunteers on a continuous basis, without the need for human action, to research monitors and clinicians.

Vital sign values are being monitored by the research team as part of monitoring the health of volunteers.

Reliability studies on the digital platform still need to be carried out at the end of the follow-up scheduled for October 23, 2022.

### Future Perspectives

At end of this study, the accuracy and precision of the collected data will be analyzed.

Once it is concluded that the data collected are reliable and have high accuracy, a larger study with volunteers across the country is planned, taking advantage of the structure already built for LIKA-Web.

A new digital solution is planned for development, with which volunteers will not only be able to access their vital signs data

but also grant permission to their caregivers, so they can also monitor their health.

In the near future, with the LIKA-Web platform, it will be possible to carry out remote monitoring of an entire community,

supporting existing primary care programs and generating alerts of altered vital signs, in order to support health care professionals.

---

## Acknowledgments

Part of the results presented in this work were obtained through the project “Predictive monitoring in person-centered care using smartwatches” financed by Samsung Eletrônica da Amazônia Ltda under the Brazilian Informatics Law no. 8,248/91

The authors would like to thank Fundação Faculdade de Medicina for funding all equipment and subsidies.

We thank the following research monitors who work daily with the volunteers: Alysson Michel Mariano, Amanda Cavalcanti de Oliveira, Andre Luis da Silva, Antenor Bispo dos Santos Silva, Fernanda Kayoko Yokota Harayashiki, Filipe Eustaquio da Silva Gomes de Oliveira, Flavio Rodrigo Cichon, Igor Staub Zago, Joana Beatriz Midoes Mariano, Mariana Cavazzoni Lima de Carvalho, Rafael Luz dos Anjos, Raquel Kaori Nagai, Ricardo Lira Araujo, and Sabrina Saemy Tome Uchiyama.

We also thank the following research administrative team for all support offered: Joao Victor Scartozzoni, Giovana Alves da Silva, Rafael Ferreira dos Reis, and Vinicius Delgado Ramos.

---

## Authors' Contributions

KJB, LRDP, FBS, and LRB conceptualized the study, curated the data, designed the methodology, supervised the study, validated the system, and visualized the data. KJB, LRDP, and FBS performed the formal analysis and the investigation; designed the software; and wrote, reviewed, and edited the manuscript. KJB and LRB acquired the funding. KJB performed project administration and provided resources. LRB reviewed and edited the manuscript.

---

## Conflicts of Interest

None declared.

---

## References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> [accessed 2022-09-06]
2. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. URL: <https://covid19.who.int/> [accessed 2022-09-06]
3. Li JO, Liu H, Ting DSJ, Jeon S, Chan RVP, Kim JE, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. *Prog Retin Eye Res* 2021 May;82:100900 [FREE Full text] [doi: [10.1016/j.preteyeres.2020.100900](https://doi.org/10.1016/j.preteyeres.2020.100900)] [Medline: [32898686](https://pubmed.ncbi.nlm.nih.gov/32898686/)]
4. Zugasti Murillo A, Aguilar Sagrañes L, Álvarez Hernández J. [Digital transformation of the relationship between industry and healthcare professionals]. *Nutr Hosp* 2022 Mar 29;38(Spec No1):14-18 [FREE Full text] [doi: [10.20960/nh.04064](https://doi.org/10.20960/nh.04064)] [Medline: [35137592](https://pubmed.ncbi.nlm.nih.gov/35137592/)]
5. Vivilaki V, Chronaki C, Barbounaki S, Petelos E. Accelerating the digital transformation of community midwifery during the COVID-19 pandemic. *Eur J Midwifery* 2021 Oct 8;5(October):44-43 [FREE Full text] [doi: [10.18332/ejm/142571](https://doi.org/10.18332/ejm/142571)] [Medline: [34708191](https://pubmed.ncbi.nlm.nih.gov/34708191/)]
6. Kakodkar P, Kaka N, Baig MN. A comprehensive literature review on the clinical presentation, and management of the pandemic coronavirus disease 2019 (COVID-19). *Cureus* 2020 Apr 06;12(4):e7560 [FREE Full text] [doi: [10.7759/cureus.7560](https://doi.org/10.7759/cureus.7560)] [Medline: [32269893](https://pubmed.ncbi.nlm.nih.gov/32269893/)]
7. Hayiroğlu M. Telemedicine: current concepts and future perceptions. *Anatol J Cardiol* 2019 Oct;22(Suppl 2):21-22. [doi: [10.14744/AnatolJCardiol.2019.12525](https://doi.org/10.14744/AnatolJCardiol.2019.12525)] [Medline: [31670712](https://pubmed.ncbi.nlm.nih.gov/31670712/)]
8. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. *Per Med* 2018 Sep;15(5):429-448 [FREE Full text] [doi: [10.2217/pme-2018-0044](https://doi.org/10.2217/pme-2018-0044)] [Medline: [30259801](https://pubmed.ncbi.nlm.nih.gov/30259801/)]
9. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nat Med* 2021 Jan;27(1):73-77. [doi: [10.1038/s41591-020-1123-x](https://doi.org/10.1038/s41591-020-1123-x)] [Medline: [33122860](https://pubmed.ncbi.nlm.nih.gov/33122860/)]
10. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. *Nat Biomed Eng* 2020 Dec;4(12):1208-1220 [FREE Full text] [doi: [10.1038/s41551-020-00640-6](https://doi.org/10.1038/s41551-020-00640-6)] [Medline: [33208926](https://pubmed.ncbi.nlm.nih.gov/33208926/)]
11. Nasarre M, Strik M, Daniel Ramirez F, Buliard S, Marchand H, Abu-Alrub S, et al. Using a smartwatch electrocardiogram to detect abnormalities associated with sudden cardiac arrest in young adults. *Europace* 2022 Mar 02;24(3):406-412. [doi: [10.1093/europace/euab192](https://doi.org/10.1093/europace/euab192)] [Medline: [34468759](https://pubmed.ncbi.nlm.nih.gov/34468759/)]

12. Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K, et al. Smartwatch algorithm for automated detection of atrial fibrillation. *J Am Coll Cardiol* 2018 May 29;71(21):2381-2388. [doi: [10.1016/j.jacc.2018.03.003](https://doi.org/10.1016/j.jacc.2018.03.003)] [Medline: [29535065](https://pubmed.ncbi.nlm.nih.gov/29535065/)]
13. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Apple Heart Study Investigators. Large-scale assessment of a smartwatch to identify atrial fibrillation. *N Engl J Med* 2019 Nov 14;381(20):1909-1917 [FREE Full text] [doi: [10.1056/NEJMoa1901183](https://doi.org/10.1056/NEJMoa1901183)] [Medline: [31722151](https://pubmed.ncbi.nlm.nih.gov/31722151/)]
14. Ford C, Xie CX, Low A, Rajakariar K, Koshy AN, Sajeev JK, et al. Comparison of 2 smart watch algorithms for detection of atrial fibrillation and the benefit of clinician interpretation: SMART WARS study. *JACC Clin Electrophysiol* 2022 Jun;8(6):782-791. [doi: [10.1016/j.jacep.2022.02.013](https://doi.org/10.1016/j.jacep.2022.02.013)] [Medline: [35738855](https://pubmed.ncbi.nlm.nih.gov/35738855/)]
15. Chae SH, Kim Y, Lee K, Park H. Development and clinical evaluation of a web-based upper limb home rehabilitation system using a smartwatch and machine learning model for chronic stroke survivors: prospective comparative study. *JMIR Mhealth Uhealth* 2020 Jul 09;8(7):e17216 [FREE Full text] [doi: [10.2196/17216](https://doi.org/10.2196/17216)] [Medline: [32480361](https://pubmed.ncbi.nlm.nih.gov/32480361/)]
16. Reeder B, David A. Health at hand: A systematic review of smart watch uses for health and wellness. *J Biomed Inform* 2016 Oct;63:269-276 [FREE Full text] [doi: [10.1016/j.jbi.2016.09.001](https://doi.org/10.1016/j.jbi.2016.09.001)] [Medline: [27612974](https://pubmed.ncbi.nlm.nih.gov/27612974/)]
17. Evaluating digital health products. UK Health Security Agency. 2021 Oct 19. URL: <https://www.gov.uk/government/collections/evaluating-digital-health-products> [accessed 2022-09-06]
18. Analysis of routinely collected data: descriptive studies. UK Health Security Agency. 2020 Jan 30. URL: <https://www.gov.uk/guidance/analysis-of-routinely-collected-data-descriptive-studies> [accessed 2022-09-06]
19. de Arriba-Pérez F, Caeiro-Rodríguez M, Santos-Gago JM. Collection and processing of data from wrist wearable devices in heterogeneous and multiple-user scenarios. *Sensors (Basel)* 2016 Sep 21;16(9):1 [FREE Full text] [doi: [10.3390/s16091538](https://doi.org/10.3390/s16091538)] [Medline: [27657081](https://pubmed.ncbi.nlm.nih.gov/27657081/)]
20. Johnson B, Clarke JM. Collecting sensitive data: the impact on researchers. *Qual Health Res* 2003 Mar 16;13(3):421-434. [doi: [10.1177/1049732302250340](https://doi.org/10.1177/1049732302250340)] [Medline: [12669341](https://pubmed.ncbi.nlm.nih.gov/12669341/)]
21. Patridge EF, Bardyn TP. Research Electronic Data Capture (REDCap). *jmla* 2018 Jan 12;106(1):1. [doi: [10.5195/jmla.2018.319](https://doi.org/10.5195/jmla.2018.319)]
22. Soni A, Ranga V. API features individualizing of web services: REST and SOAP. *IJITEE* 2019 Aug 23;8(9S):664-671. [doi: [10.35940/ijitee.i1107.0789s19](https://doi.org/10.35940/ijitee.i1107.0789s19)]
23. Patel A, Reddy V, Araujo J. *Physiology, Sleep Stages*. Treasure Island, FL: StatPearls Publishing; 2022.
24. Samsung Eletronica da Amazonia Ltda. ANVISA. URL: <https://consultas.anvisa.gov.br/#/saude/25351682843202013/> [accessed 2022-09-06]
25. Torp KD, Modi P, Simon LV. *Pulse Oximetry*. Treasure Island, FL: StatPearls Publishing; 2022.
26. Hafeez Y, Grossman SA. *Sinus Bradycardia*. Treasure Island, FL: StatPearls Publishing; 2022.
27. Jung H, Kim D, Lee W, Seo H, Seo J, Choi J, et al. Performance evaluation of a wrist-worn reflectance pulse oximeter during sleep. *Sleep Health* 2022 Jul 08;706 [FREE Full text] [doi: [10.1016/j.sleh.2022.04.003](https://doi.org/10.1016/j.sleh.2022.04.003)] [Medline: [35817700](https://pubmed.ncbi.nlm.nih.gov/35817700/)]
28. Vijayan V, Connolly JP, Condell J, McKelvey N, Gardiner P. Review of wearable devices and data collection considerations for connected health. *Sensors (Basel)* 2021 Aug 19;21(16):1 [FREE Full text] [doi: [10.3390/s21165589](https://doi.org/10.3390/s21165589)] [Medline: [34451032](https://pubmed.ncbi.nlm.nih.gov/34451032/)]
29. Bin K, Alves P, Costa R, Eiras P, Araujo L, Pereira A. Digital primary health care, building a pilot model and evaluating patient satisfaction in Santarem, Amazon. *J Med Internet Res (forthcoming)* 2022:1. [doi: [10.2196/preprints.39034](https://doi.org/10.2196/preprints.39034)]

## Abbreviations

**cBPM:** continuous beats per minute

**HCFMUSP:** Hospital das Clínicas of The Faculdade de Medicina da Universidade de São Paulo

**mHealth:** mobile health

**REDCap:** Research Electronic Data Capture

**REM:** rapid eye movement

**Q1:** lower quartile

**Q2:** median quartile

**Q3:** upper quartile

*Edited by A Mavragani; submitted 22.06.22; peer-reviewed by Y Kim, F Balzer; comments to author 14.07.22; revised version received 01.08.22; accepted 23.08.22; published 15.09.22.*

*Please cite as:*

*Bin KJ, De Pretto LR, Sanchez FB, Battistella LR*

*Digital Platform to Continuously Monitor Patients Using a Smartwatch: Preliminary Report*

*JMIR Form Res 2022;6(9):e40468*

*URL: <https://formative.jmir.org/2022/9/e40468>*

*doi: [10.2196/40468](https://doi.org/10.2196/40468)*

*PMID: [36107471](https://pubmed.ncbi.nlm.nih.gov/36107471/)*

©Kaio Jia Bin, Lucas Ramos De Pretto, Fabio Beltrame Sanchez, Linamara Rizzo Battistella. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 15.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Using Wake-Up Tasks for Morning Behavior Change: Development and Usability Study

Kyue Taek Oh<sup>1</sup>, MS; Jisu Ko<sup>2</sup>, BS; Jaemyung Shin<sup>3</sup>, MS; Minsam Ko<sup>1</sup>, PhD

<sup>1</sup>Department of Human-Computer Interaction, Hanyang University, Ansan, Republic of Korea

<sup>2</sup>Department of Applied Artificial Intelligence, Hanyang University, Ansan, Republic of Korea

<sup>3</sup>Delightroom, Seoul, Republic of Korea

**Corresponding Author:**

Minsam Ko, PhD

Department of Human-Computer Interaction

Hanyang University

Ansan

Republic of Korea

Phone: 82 314001071

Email: [minsam@hanyang.ac.kr](mailto:minsam@hanyang.ac.kr)

**Related Article:**

This is a corrected version. See correction statement: <https://formative.jmir.org/2022/10/e42926>

## Abstract

**Background:** Early morning behaviors between waking up and beginning daily work can develop into productive habits. However, sleep inertia limits the level of human ability immediately after waking, lowering a person's motivation and available time for productive morning behavior.

**Objective:** This study explores a design for morning behavior change using a wake-up task, a simple assignment the user needs to finish before alarm dismissal. Specifically, we set two research objectives: (1) exploring key factors that relate to morning behavior performance, including the use of wake-up tasks in an alarm app and (2) understanding the general practice of affecting morning behavior change by implementing wake-up tasks.

**Methods:** We designed and implemented an apparatus that provides wake-up task alarms and facilities for squat exercises. We recruited 36 participants to perform squat exercises in the early morning using the wake-up tasks for 2 weeks. First, we conducted a generalized estimating equation (GEE) analysis for the first research objective. Next, we conducted a thematic analysis of the postsurvey answers to identify key themes about morning behavior change with the wake-up tasks for the second objective.

**Results:** The use of wake-up tasks was significantly associated with both the completion of the target behavior (math task:  $P=.005$ ; picture task:  $P<.001$ ) and the elapsed time (picture task:  $P=.08$ ); the time to alarm dismissal was significantly related to the elapsed time to completion ( $P<.001$ ). Moreover, the theory of planned behavior (TPB) variables, common factors for behavior change, were significant, but their magnitudes and directions differed slightly from the other domains. Furthermore, the survey results reveal how the participants used the wake-up tasks and why they were effective for morning behavior performance.

**Conclusions:** The results reveal the effectiveness of wake-up tasks in accomplishing the target morning behavior and address key factors for morning behavior change, such as (1) waking up on time, (2) escaping from sleep inertia, and (3) quickly starting the desired target behavior.

(*JMIR Form Res* 2022;6(9):e39497) doi:[10.2196/39497](https://doi.org/10.2196/39497)

**KEYWORDS**

health app design; morning behavior change; wake-up task; mobile alarm; productivity

## Introduction

### Background

There has been rising interest in spending early morning time productively, with a view that early morning behaviors between waking up and the starting of daily work can develop into productive habits (called miracle morning) [1,2]. These habits improve health and quality of life [3]. For example, morning exercise is effective in increasing physical performance, such as muscular strength [4], anaerobic power [5], and endurance [6], while improving blood sugar [7] and hormone levels [8]. Moreover, early morning emotions can influence feelings throughout the day [9].

Many researchers have studied human behavior change [10]. For example, a conscious goal can induce the anticipated behavior [11]. Locke and Latham [12] discovered a linear relationship in which the appropriate difficulty of the goal produces a high level of effort and performance. Fogg [13,14] explained the initiation of the target behavior by 3 factors (motivation, ability, and trigger) and presented the behavior grid, which categorizes behavior change into 35 cases according to behavior type and maintenance degree. Prochaska et al [15,16] described 6 phases of behavior change (precontemplation, contemplation, preparation, action, maintenance, and termination). Such theories can provide valuable insights for developing intervention methods but are often insufficient to explain specific behavior changes in a health domain [17,18].

Under these behavior theories, studies have sought to develop computational supports for behavior change and persuasion [19-21]. Oinas-Kukkonen et al [21] classified persuasive system principles into 4 categories (primary task, dialog, system credibility, and social support). For example, tunneling can facilitate the target behavior by allowing the process to proceed without interruption. Fogg [19] proposed social cues for the persuasive system as a social actor. Oulasvirta et al [22] investigated smartphone usage from the perspective of habit formation. Consolvo et al [23] studied behavior change components by designing a mobile app that uses social supports to foster healthy behavior. There have also been studies on the impact of mobile tools on behavior change. For example, ReVibe [24] uses contextual information automatically collected from phones or sensors to improve momentary ecological assessments. FoodPrint [25] is a photo-based food diary that helps patients and health experts exchange knowledge and focus on collaboration goals. Costa et al [26] proposed a method for regulating emotions through haptic feedback with a smartwatch.

However, it is necessary to consider contextually unique characteristics for morning behavior change that differ from the prior studies in general contexts. For example, physical and cognitive performance decreased in the sleep inertia state. This is especially true of a person's decision-making ability [27,28] and indicates that prior behavior change mechanisms may not work effectively in the morning. Thus, this study focuses on morning behavior change and attempts to understand design implications considering the contextual characteristics of morning hours.

We especially note that sleep inertia [29], which refers to the phenomenon that occurs after waking up in which people's abilities are significantly lower than those under their normal condition, characterizes early morning behavior. Several studies have investigated physical and mental states when experiencing sleep inertia to understand the contextual characteristics of sleep inertia [28,30,31]. For example, it takes more time for someone experiencing sleep inertia to perform arithmetic calculations [32,33] and recognize objects [34,35]. Although sleep inertia usually vanishes naturally within some minutes of waking, its duration varies depending on the individual and context [34,35]. Efficiently completing target morning behaviors at a predictable time is essential because a prescheduled daily routine begins sooner (eg, going to work). Therefore, sleep inertia would be a critical factor in a person's morning behavior performance.

Various studies aimed to determine how to escape sleep inertia quickly and effectively [36]. For example, McFarlane et al [37-39] explored auditory factors, such as alarm sounds, to decrease the influence of sleep inertia, and Hilditch et al [40] attempted to induce arousal using light of specific frequencies. Additionally, physical activity has been addressed as the primary countermeasure to sleep inertia. For example, Kaplan et al [41] proposed a method to reduce the sleep inertia duration and severity by performing a series of physical activity routines. Kovac et al [42] also measured the degree of sleepiness after the experimental participants performed the postwaking cycle and sprint.

Our study aligns with these studies in terms of helping escape sleep inertia. Although these studies focused on developing methods to escape from sleep inertia quickly and effectively, our study aims to help users maintain specific behaviors against sleep inertia, leading to morning behavior changes by associating them with a mobile alarm.

### Objectives

This study proposes a mechanism for morning behavior change based on wake-up tasks in a mobile alarm. A wake-up task is a simple assignment that the user needs to finish first to dismiss a ringing alarm, such as taking a picture of a particular object or solving math problems. Its effectiveness in waking up on time has been analyzed [43,44]. In this study, we designed and implemented a mobile alarm app based on a wake-up task to facilitate a target behavior in the morning. We also conducted formative research to understand morning behavior change by wake-up tasks, considering the following objectives: (1) exploring key factors that relate to morning behavior performance, including the use of wake-up tasks and (2) understanding the general practice of effecting morning behavior change by implementing wake-up tasks.

## Methods

### Apparatus

#### Design Requirements

We conducted a preliminary survey to understand the practice and intention of early morning behaviors. Our survey included several open-ended questions organized into 3 sections. The

first section includes 2 open-ended questions about respondents' usual morning behaviors and intentions toward productive morning behavior. The second section asks respondents to describe their previous efforts and strategies for productive morning behaviors. Finally, the survey asked all respondents to evaluate their past morning behavior and describe any difficulties they encountered. The survey was conducted through the internet and posted on social networking services in South Korea. The survey was answered by 40 participants (age: mean 30.32, SD 3.87 years), of which 17 were men (age: mean 31.64, SD 4.16 years) and 23 were women (age: mean 29.34, SD 3.04 years). The respondents were compensated with an electronic voucher worth approximately US \$4. We conducted a thematic analysis following the guideline proposed by Braun and Clarke [45]. The transcripts of the responses were reread by 2 researchers, and they generated initial codes. Next, they examined the similarity between codes to define the final themes. The survey respondents are denoted by R1-R40.

First, the results show usual morning behaviors immediately after waking up. For example, 60% (24/40) of the respondents said they usually check their smartphones after waking up. Specifically, among respondents who said they usually check their smartphones, 41.6% (10/24) use social networking service or messenger apps, another 41.6% (10/24) check the time, and 16.6% (4/24) read the news on the internet. Other respondents (16/40, 40%) also reported other activities, including drinking water, going to the bathroom, making the bed, and stretching.

Furthermore, 87.5% (35/40) of respondents commented that they wish to perform more productive behaviors in the morning. For example, 1 respondent said, "I can do simple exercise at least. I think it will be good for my health" (R10). Another respondent mentioned, "I think I can start the day in a good mood if I am more productive in that hours" (R25). Some respondents also addressed the benefits of using early morning time for productive purposes. For example, a participant said, "Taking time to do something in the evening is difficult. But, in the morning, as long as I get up early, I can have extra time to do meaningful behavior, such as stretching" (R12). On the other hand, 5 respondents showed negative attitudes, mainly because they had been too busy in the morning. For example, a participant stated, "I'm too busy getting ready for work, so I don't have much time to do productive activities" (R32).

Additionally, 75% (30/40) of the respondents had previously tried productive morning behavior. Targeted morning behaviors

were mostly related to physical activity, such as stretching or simple exercising (eg, squats). Other activities like reading, writing, studying, meditation, and praying were also mentioned. These respondents explained their strategies for maintaining their desired behavior and getting up earlier is the most common strategy. Respondents also said they tended to put considerable effort into remembering the target behavior, such as forcing themselves to perform the behavior in a group setting. Next, several respondents said they had attempted to find suitable behaviors for the available morning time and had modified the goal based on the situation and performance. For example, some respondents thoroughly planned their morning time to ensure that the target behavior finished in time, whereas others attempted to find simple stretching exercises.

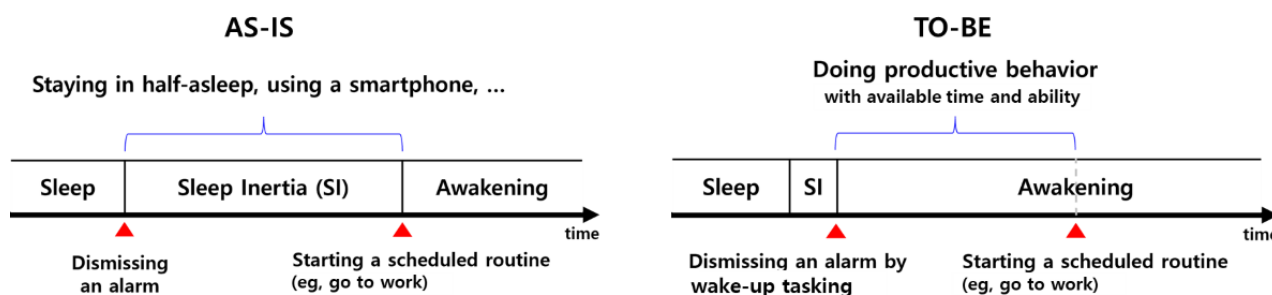
However, maintaining the target behavior every morning was challenging. Most responses were related to a lower level of awakening, with 82.5% (33/40) of respondents stating they often woke up late or stayed half-asleep excessively. There were also other awakening-related comments, such as demotivation. For example, "When I got up in the morning, I did not want to do anything and hesitated to start exercising, although I planned it" (R4).

**Design Implications: Winding Up and Nudging by Wake-Up Tasks**

Most failures affecting the targeted morning behavior are related to the characteristics of morning hours, such as limited time and respondents' lowered ability due to sleep inertia. Therefore, 2 requirements for morning behavior support are proposed. First, it is critical to ensure that the user wakes up on time and recovers to the normal state as soon as possible. Second, it is necessary to create a bridge between waking up and starting the target behavior to prevent forgetting to perform it and avoid distractions.

This study proposes a new mechanism for morning behavior change by implementing a wake-up task, as shown in Figure 1. The wake-up task means a user must perform a simple task to dismiss a ringing alarm (eg, taking a picture or solving math problems) [44]. Completing a wake-up task can help users recover their normal ability quickly and have sufficient time for the target behavior. Specifically, the proposed mechanism works in two stages: (1) winding up to escape from sleep inertia by performing a wake-up task and (2) nudging to induce the target behavior after completion.

Figure 1. Proposed mechanism for morning behavior change by wake-up tasking.



## Implementation

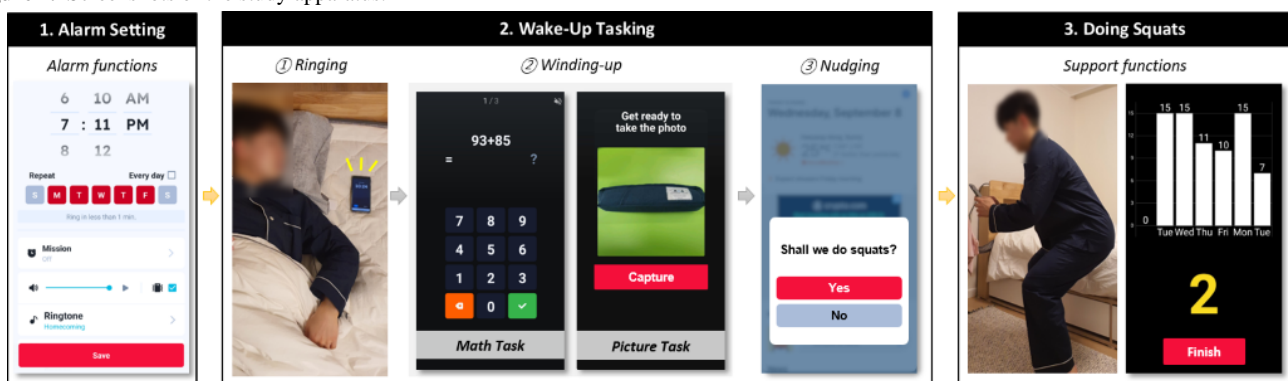
We implemented an alarm app that provides task-based alarm dismissal, as shown in Figure 2. The app provides three alarm dismissal methods: a conventional method as a baseline (ie, pressing a button) and two wake-up tasks (taking a picture and solving math problems). These are denoted as `non_task`, `picture_task`, and `math_task`, respectively. In this study, a squat exercise was selected as a target behavior. This is because simple exercising was the most preferred morning behavior, according to the preliminary study. Furthermore, doing squats is relatively easy and does not require additional equipment. The app functions as follows:

First, the app provides standard alarm functions, such as controlling the alarm sound, switching the alarm to vibrate mode, and selecting a ringtone from built-in sound sources. Notably, the user must choose an alarm dismissal method from among `non_task`, `math_task`, and `picture_task`. Furthermore, `picture_task` and `math_task` are representative tasks that require physical and cognitive abilities, respectively. For setting the `picture_task`, the user must input a reference image for later authentication. The app encourages users to select an image taken far away from their bed as the reference image. For the

`math_task`, the user specifies the number of problems and the difficulty level. There are 6 levels of difficulty. The number of digits a user needs to calculate increases as the difficulty increases, and multiplication operations are included from level 4. The user needs to solve 3 problems by default but is allowed to change the number.

Next, when the alarm goes off, the user must complete the task required by each dismissal method; this is the winding-up stage to help escape from sleep inertia. `non_task` is a conventional dismissal method in many alarm apps. With `non_task`, the user can dismiss an alarm by simply pressing the dismissal button. `math_task` requires performing arithmetic calculations for the alarm dismissal, and the alarm will keep ringing until the user correctly answers the problems. With `picture_task`, the user must take the same picture as the registered reference image. The app determines whether the input and reference images are the same based on their image features (ie, color, shape) extracted by Android OpenCV [46]. After dismissing the alarm, the nudging stage begins with a dialog popup asking whether the user will start the target behavior immediately. If the user accepts it, the app displays functions to help the user perform the required squat exercises.

Figure 2. Screenshots of the study apparatus.



Finally, the user begins to perform the target behavior with support functions provided by the app. The users can view their performance records for the last 7 days. During the squat exercise, the app also counts the number of squat reps performed by the user and displays them on the smartphone screen. When the user performs a squat by grabbing the smartphone and looking at the screen, the app recognizes the up-and-down motion using the internal inertial measurement unit sensor values for the past 2 seconds. We built a simple deep learning model based on long short-term memory units using TensorFlow, and the trained model was converted into the TensorFlow-Lite format, which enables it to run on a mobile device. When the user finishes the exercise by pressing the “complete” button, the app asks the user to input the exact number of the reps to correct the miscounted cases. The main experiment with 36 participants discovered no significant difference between the automatically counted number of squats and the number the user inputted (auto: mean 13.5, SD 6.8; manual: mean 14.6, SD 4.4). Their Pearson correlation coefficient,  $r=0.636$ , and the mean absolute error (MAE) was 2.65, indicating 2 or 3 miscounted reps. This tendency became stronger when 3 users who had consistently reported frequent miscounts from the

beginning (possibly due to device or operating system issues) were excluded (MAE=1.99,  $r=0.717$ ).

## Participants

A between-group study asked users to try to do squat exercises with the apparatus in the early morning for 2 weeks. The target behavior for this study was performing 15 squats 10 minutes after the target time was set for the morning alarm. We employed generalized estimating equation (GEE) modeling in this study to measure continuous and repetitive behavioral changes. We investigated previous studies that used the GEE model as a statistical method and confirmed the sample size of this study by referring to the experimental configuration of previous studies [47-49]. Therefore, we recruited 36 participants (21 males and 15 females) who can use mobile alarms and perform squats every morning through social networking service posting. The average age of these users was 27.5 years (SD 7.68 years), and they were mostly office workers (20/36, 55.6%) and students (16/36, 44.4%). We denote these participants by P1-P36.

## Procedure

First, the participants were asked to respond to a presurvey. The presurvey asked about their demographics and usual wake-up time on a weekday. Notably, the survey included the theory of planned behavior (TPB) [50], which consists of 20 items on a 7-point Likert scale to assess traditional behavior factors (ie, intention, attitude, subjective norm, and control). Next, participants were divided into 3 groups by assigning individuals to 1 of the dismissal methods (non\_task: 12, picture\_task: 12, and math\_task: 12). We confirmed that all variables measured by the presurvey (ie, age, gender, and TPB score) did not significantly differ among the groups. The detailed results of the presurvey are described in [Multimedia Appendix 1](#).

Next, the 2-week experiment began, with each participant instructed to set a morning alarm using the implemented app with the assigned dismissal method. Participants were encouraged to set their alarms 5 minutes earlier to improve morning behavior. Each alarm was individually checked to ensure that it worked as intended; all users were using the dismissal method, and they performed their squat exercises with the app's counter. Finally, participants were asked to perform 15 squats every weekday morning for 2 weeks voluntarily. We emphasized that they could skip or reject the behavior of their choosing, and we assured them that they would receive the same compensation regardless of their performance.

A postsurvey was conducted once the experiment was completed to understand detailed contexts about early morning behavior with wake-up tasks. This survey first inquired about their general usage and experience of the app, including helpful functions and failure cases. The next section of the survey questioned their alarm settings and measured the workloads of dismissing an alarm early in the morning by the NASA-Task Load Index (NASA-TLX) questionnaire [51]. We used the original NASA-TLX questionnaire in a graphic form that consists of rating bars [51]. The rating bar provides effective assistance for experimental respondents to understand and respond to the questionnaire intuitively, so we judged that the influence of linguistic factors on the experimental results was insignificant. Finally, the survey asked users to explain changes in their daily lives after the experiment.

## Analysis Methods

### *Quantifying Morning Behavior Performance*

Behavior performance was quantified from two perspectives: (1) success rate and (2) time elapsed for users to start their first squat in a successful trial. The target behavior of the day was regarded as a success if the user completed 15 squats. This was denoted as a binomial variable, success, which had a value of 1 when it was a success and 0 otherwise. If the user completed the squats 15 times within 10 minutes after the morning alarm fired, the case was called early\_success (note that success is a variable that includes early\_success).

Unlike other domains of behavior change, early starts can be important in the morning hours because frequent morning delays cause failures by running out of available morning time. As a result, the target behavior was evaluated by measuring the elapsed time to starting the first squat in a successful trial after

the alarm rang (ET\_ring2success). The time between alarm ringing and dismissal (ET\_ring2dismiss) was also measured, as was the time between when users dismissed the alarm and when they started their first squat in a successful trial (ET\_dismiss2success).

### *Exploring Factors Affecting Morning Behavior Performance*

Next, a GEE [52] analysis was performed to determine whether and how the wake-up task and alarm usage affected morning behavior performance. This study used repeated measures (eg, success, elapsed time) for the same participant. Therefore, the GEE method, widely used to handle repetitive measurements or for clustering data, was adopted [53,54]. The GEE analysis is also suitable for this study (compared with other repeated-measures analyses, including repeated-measures ANOVA) because it can use more than 2 diverse predictors. TPB behavior factors could explain behavior change better by cooperating with the alarm and wake-up task usage. Specifically, 2 dependent variables related to the performance were used, namely early\_success and ET\_ring2success. According to the variable type, a binomial GEE model was developed using the logit link function for early\_success and another with Gaussian estimation for ET\_ring2success. In total, there were 9 independent variables. The 3 groups representing the dismissal methods were represented by 2 dummy variables (eg, picture\_task=1 and others=0; math\_task=1 and others=0). Two variables related to alarm usage were considered: the time the alarm fires (ringing\_time) and ET\_ring2dismiss. Then, four subscales in the TPB questionnaire were used: (1) intention (TPB\_intention), (2) attitude (TPB\_attitude), (3) subjective norm (TPB\_subjective\_norm), and (4) control (TPB\_control). Finally, we used the elapsed days (ET\_days) as the last independent variable to handle repeatedly measured data.

### *Analyzing the Practice of Early Morning Behavior With Wake-Up Tasks*

Finally, we aimed to understand the practice of early morning behavior with wake-up tasks based on the app's usage logs and postsurvey responses. We conducted both quantitative and qualitative analyses. First, the NASA-TLX score on wake-up tasks was aggregated by following its original guideline [51], which determines weights for workload dimensions and calculates their weighted average. We conducted a 1-way ANOVA test to compare the NASA-TLX scores among the 3 groups, and a Tukey Honestly Significant Difference test was performed as the posthoc test. Second, similar to the preliminary study, 2 researchers collaboratively performed thematic analysis on postsurvey responses.

### **Ethical Considerations**

All subjects participated voluntarily and received US \$43 each as compensation. All the study data are deidentified. This study has received an exemption from the Institutional Review Board of the Hanyang University (HYUIRB-202205-011).

## Results

### User Statistics

#### Success Rate

Figure 3 depicts the success rate. The picture\_task group achieved the highest early\_success rate (94.2%), followed by the math\_task group (87.5%). This indicates that these user groups maintained the target behavior almost every weekday for 2 weeks. However, the participants in the non\_task group, which used the conventional dismissal method, finished their

daily squats mission only 75.8% of the time during the same period. A similar tendency occurred when adding successes to the quantity. The picture\_task group (97.5%) and math\_task group (94.2%) showed a success rate higher than 90% when the time limit was ignored.

The non\_task group's success rate also increased with successes, but it remained the lowest (80.8%). As shown in Figure 4, we could not find a clear tendency over time for all groups. However, the non\_task (SD 11.64) group showed a relatively higher variability in the daily success rate than the picture\_task (SD 6.72) group and the math\_task (SD 6.01) group.

Figure 3. Average success rate. SD: standard deviation.

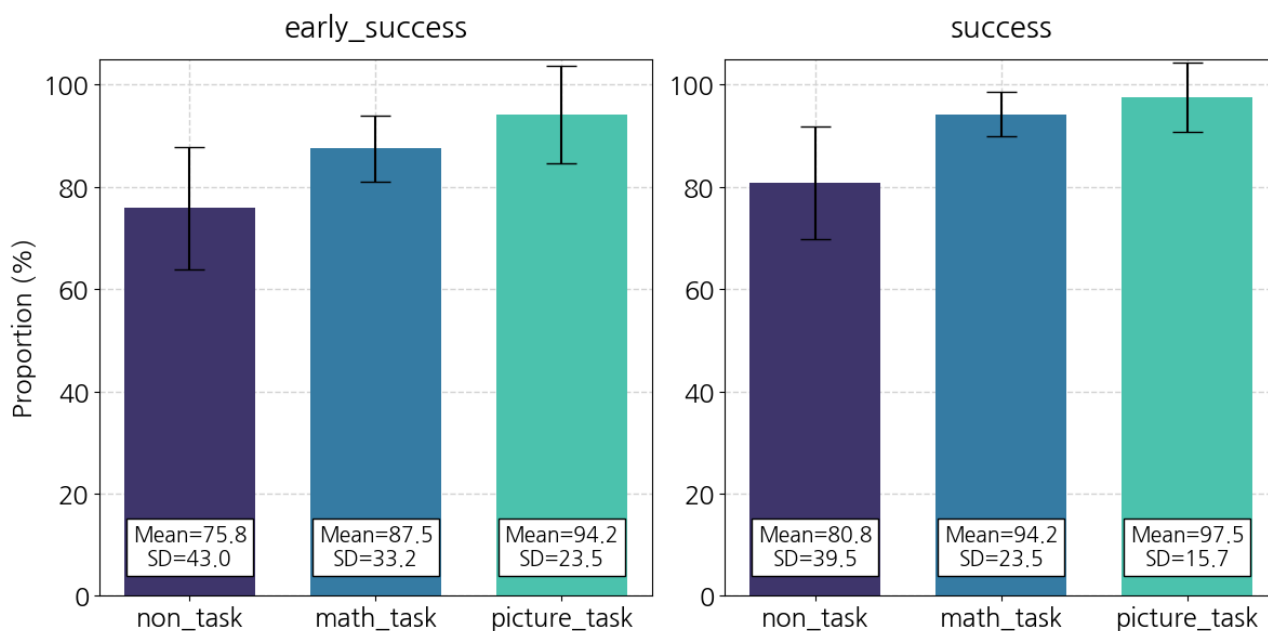
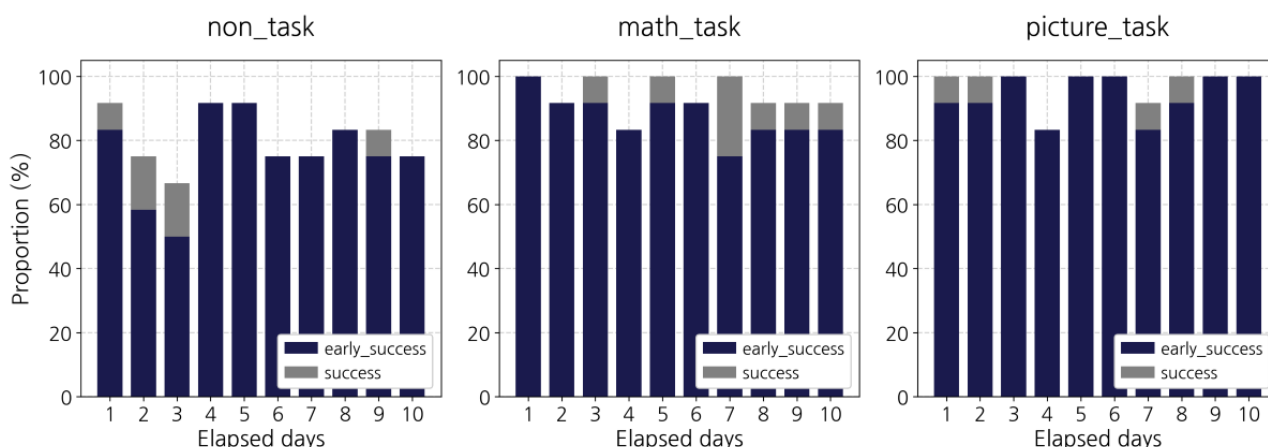


Figure 4. Daily success rates.



#### Elapsed Time to Start the First Squat in a Successful Trial

Figure 5 illustrates the timeline from alarm ringing to when the users started their first squat in a successful trial for each user group. Note that the figure presents only the trials in which the user successfully performed squats at least 15 times. We discovered that the picture\_task group had the shortest time to success after the alarm rang (84.4 seconds). The math\_task

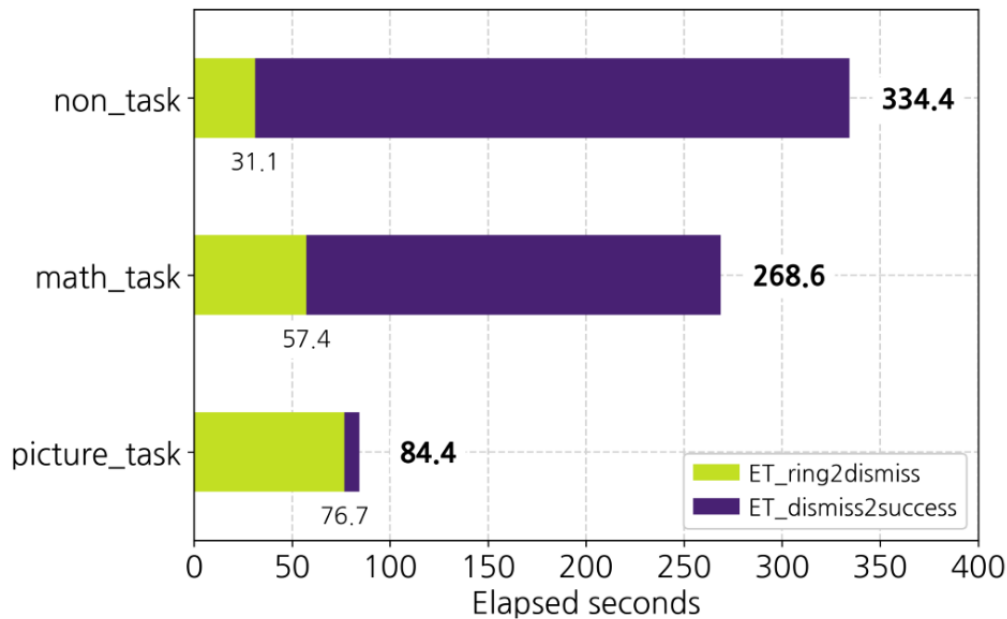
group (268.6 seconds) came in second, followed by the non\_task group (334.4 seconds). The math\_task and picture\_task groups took 19-45 seconds longer than the non\_task group to dismiss their alarms. This is probably because the user was required to move to a specified location to capture the preregistered image or perform the given arithmetic operations.

Nevertheless, the elapsed time to start the first squat in a successful trial tended to be much shorter in the math\_task and

picture\_task groups than in the non\_task group. In particular, the picture\_task group started squats in a very short time of 7.7 seconds after dismissing the alarm. On the other hand, the

non\_task group's ET\_dismiss2success time is relatively longer than that of the other groups, indicating that the non\_task group often delayed the target behavior after dismissing the alarm.

**Figure 5.** Average elapsed seconds to begin (in successful trials).



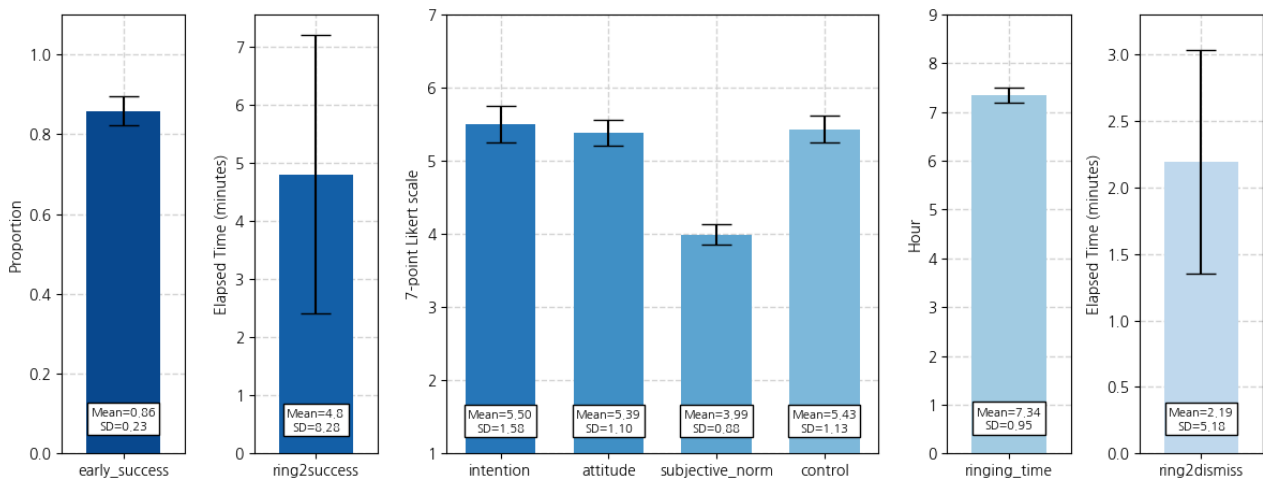
**Outcome 1: Factors Affecting Morning Behavior Performance**

Figure 6 shows the descriptive statistics of 8 variables. On average, 86% of the participants completed the daily target behavior successfully. The average elapsed time to start the first squat in a successful trial was approximately 4.5 minutes. Both intention and attitude scores were greater than 5 out of 7, indicating that participants tended to be interested in (and had a positive attitude toward) doing squats in the early morning. Participants also seemed to have confidence in doing squats in the early morning. The TPB\_subjective\_norm score was relatively lower than the others, suggesting that the participants tended not to feel external pressures or expectations about this

behavior. Finally, according to ringing\_time and ET\_ring2dismiss, participants mostly set the wake-up alarm to go off at around 7 AM and spent 2.2 minutes on average, dismissing the alarm after it rang.

Table 1 displays each independent variable's coefficient in the GEE analysis results. First, we discovered that the alarm dismissal method significantly affected the achievement of early\_success. The picture\_task was significantly related to the success of morning behavior compared with the non\_task (odds ratio [OR] 47.1, 95% CI 2.154-5.554). Similarly, the odds of the math\_task group were 3.99 times higher than that of the non\_task group (95% CI 0.423-2.347). We also discovered that the odds of success increased more as the user dismissed the alarm earlier (OR 0.652, 95% CI -0.685 to -0.172).

**Figure 6.** Statistics of the variables for generalized estimation equation analysis.



**Table 1.** Results of the generalized estimating equation analysis.

Variables	early_success				ET_ring2success			
	$\beta$	SE	<i>P</i>	95% CI	$\beta$	SE	<i>P</i>	95% CI
<b>Wake-up task</b>								
math_task	1.384	0.491	.005	0.423 to 2.347	1.282	2.366	.59	-3.355 to 5.921
picture_task	3.854	0.868	<.001	2.154 to 5.554	-4.052	2.288	.08	-8.537 to 0.431
non_task (reference category)	N/A <sup>a</sup>	N/A	N/A	N/A	N/A	N/A	N/A	N/A
<b>Alarm usage</b>								
ringing_time	-0.104	0.229	.65	-0.553 to 0.344	-1.473	1.423	.30	-4.263 to 1.316
ET_ring2dismiss	-0.428	0.131	.001	-0.685 to -0.172	1.006	0.010	<.001	0.988 to 1.025
<b>Theory of planned behavior (TPB)</b>								
TPB_intention	0.605	0.321	.06	-0.024 to 1.234	0.485	0.812	.55	-1.107 to 2.077
TPB_attitude	0.408	0.566	.47	-0.702 to 1.518	-3.763	2.420	.12	-8.506 to 0.979
TPB_subjective_norm	-1.126	0.393	.004	-1.898 to -0.356	4.143	1.911	.03	0.397 to 7.890
TPB_control	-0.218	0.388	.57	-0.979 to 0.542	0.151	1.387	.91	-2.568 to 2.870
<b>Days</b>								
ET_days	-0.013	0.051	.80	-0.114 to 0.087	-0.146	0.288	.61	-0.711 to 0.418

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

Next, as shown in Table 1, there were significant factors influencing the elapsed time to success. The subjective norm in TPB was significantly and positively correlated with the elapsed time, indicating that those who feel less pressured tend to begin their squat exercises more quickly. The elapsed time to success tended to be significantly shorter, as the user dismissed the alarm earlier. Furthermore, a marginally significant relationship between picture\_task and the elapsed time to success was found, indicating that the user tended to finish the target behavior more quickly when using the picture\_task. There was no significant relationship between the other variables and the elapsed time to success.

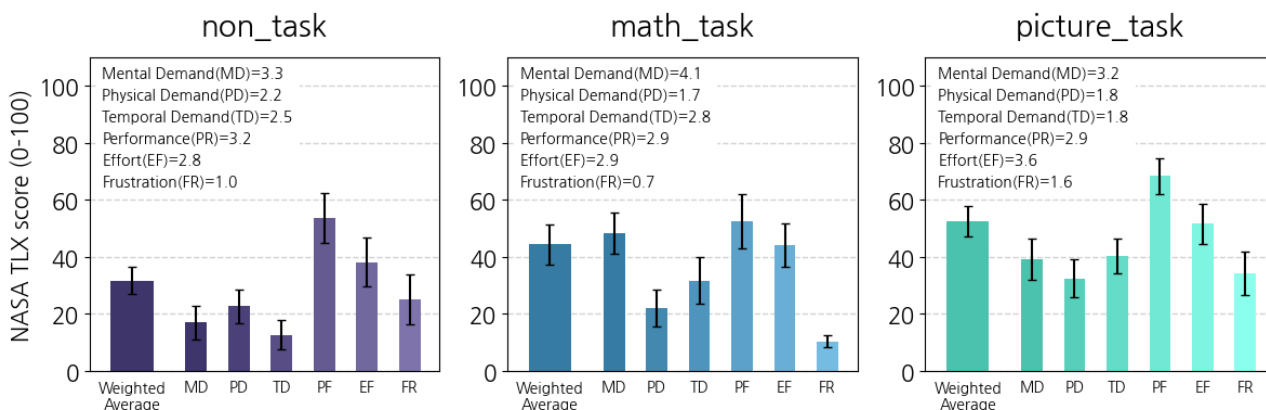
**Outcome 2: General Practice of Morning Behavior Change by Wake-Up Tasks**

**Quantitative Analysis Results**

As shown in Figure 7, the total TLX scores of all 3 groups were under 50, indicating an insignificant psychological burden by

all methods. The non\_task group’s workload was the lowest, followed by the picture\_task group and math\_task group. We found that the difference in the scores was closely significant and large among the 3 groups’ TLX scores ( $F_{2,33}=3.106, P=.06$ , and  $\eta^2=.158$ ), and the posthoc testing results reveal that there was the significant difference between the non\_task group and the picture\_task group (adjusted  $P=.048$  and  $d=1.164$ ). For the 6 subscales, the performance was the highest for all 3 groups, possibly because this study’s tasks involved waking up on time in the morning, which may have satisfied the participants. We also discovered that the type of wake-up tasks appeared to be related to each subscale. For example, the math\_task group users tended to perceive a relatively higher mental demand. On the other hand, the picture\_task group felt that their task was physically demanding and frustrating.

**Figure 7.** Workloads of alarm dismissal methods (NASA-Task Load Index).





## Qualitative Analysis Results

### Morning Behavior Supports

The participants in this study indicated that they appreciated app functions that supported morning behaviors. First, waking up on time with wake-up tasks was cited as helpful. For example, 1 participant commented, *“Sometimes I just turn off the alarm and lie down again. But, solving a math problem after the alarm fires helped me wake-up easily”* (P34). Another participant, who was a part of the picture\_task group, said, *“I had to get up and adjust the position and angle of the target object. I think this makes me wake-up surely, and it became easier to do squats”* (P15). Furthermore, most failure cases were related to waking up late. For example, 1 participant mentioned, *“I turned off the alarm and just tried to sleep a little more. But when I woke up later, over 20 min had passed”* (P7). Second, more than half of the participants (22/36, 61.1%) mentioned the nudging dialog as a useful app function that supported their successful behavior change. It reminded and motivated them to perform the target behavior, with 1 participant saying, *“I often forgot about squats when I just had woke up. But, the dialog message motivated me to start squats each time”* (P18). Similarly, another participant said, *“Even when I finished taking a picture, I sometimes wanted to reject to do squats. But, I felt motivated after seeing the dialog message, and then that button led me to do squats”* (P16). Finally, having the app automatically count the squat reps was cited as helpful by users. One participant said, *“It was convenient that the app counts the number of squats, and I didn’t need to pay much attention to count it”* (P36). Some participants also mentioned that visualizing daily performance encouraged them to keep doing the squats. For example, 1 participant commented, *“I was motivated to do harder when I saw the date that I didn’t complete the squats”* (P27).

### Alarm Setting

We investigated the participants’ alarm setting and their perceived workload. All participants set their alarms for around 7 AM and rarely changed the alarm time. Most participants set the alarm as usual or 5-10 minutes earlier. A participant said, *“I woke up a little earlier to perform the task and squats so that I could keep the same routine as before, such as having breakfast and going to work on time”* (P26). Some participants reported the need to change the alarm time due to a schedule change, but this was not frequent. Next, we analyzed the math\_task group’s usage. Considering individual differences in arithmetic operations, we initially allowed participants with the math\_task setting to adjust the difficulty level and the number of problems. In the math\_task group, 75% (9/12) of the users performed 2 to 3 arithmetic operations that involved adding and subtracting 2 or more numbers of 2 digits. The others preferred more difficult arithmetic operations (ie, multiplication) or more problems. One participant said, *“The example problems were too easy for me, so I increased the difficulty level.”* The picture\_task group users usually selected their reference image from between 3 and 5 meters outside their bedroom, such as in their living room or bathroom. The most preferred object for the reference image was a wash basin.

### Positive Changes in Daily Life

First, several participants noted an improvement in their awakening level. For example, a participant said, *“I have felt more awake in the morning than usual”* (P19). Other participants reported a significant reduction in their half-asleep state in bed. One said, *“I was able to spend less time lying down and dawdling in a bed”* (P29). Second, some participants commented that the app helped them maintain a regular and productive daily routine. These responses were mostly related to waking up on time. By waking up regularly at a target time, participants felt they had more time to spare in the morning and that their daily routines were more consistent. A participant said, *“I liked I could get more time for getting ready for work or taking care of children when I regularly woke up early”* (P25). Another participant stated, *“Usually, I woke up irregularly according to the schedule on that day. But I could live a more consistent life in the past two weeks”* (P32). Finally, some participants provided meaningful comments saying that they started their day in a more positive mood. One participant stated, *“I felt refreshed in the morning as I started a day with a simple exercise”* (P35).

## Discussion

### Principal Results

#### Effectiveness of Wake-Up Tasks for Morning Behavior Change

The results of this study indicate that wake-up tasks can help users effectively use available morning time and maintain their target behavior. Wake-up tasks were significantly and positively associated with target behavior success in the GEE analysis results. The wake-up task groups’ elapsed times to success tended not to be much longer than those of the non\_task groups, even though they resulted in a 30-40 second delay in alarm dismissal. When wake-up tasks were completed, the morning time saved relieved the burden of performing the targeted morning behavior, especially for those who had busy mornings (see the comment by R32 for reference). Additionally, participants appreciated the key components of the alarm app developed for this study (ie, alarm, wake-up tasking, and nudging dialog). The mobile alarm helped participants wake up timeously, and the wake-up task after that brought about a winding-up effect by bothering them. Finally, the nudging dialog within the app seemed helpful in getting participants to start their squat exercise routine quickly.

These results are consistent with Fogg’s MAP (motivation, ability, and prompt) theory [13]. The MAP theory outlines 3 factors (motivation, ability, and prompts) to explain behavioral results. In the morning hours, the motivation and ability for the target behavior are likely lower than usual, which can disrupt the behavior’s initiation. Thus, the mobile alarm with wake-up tasks can play a role as an effective prompt for morning behavior by efficiently recovering the normal state of motivation and ability through awakening, warming up, and motivating the user.

Furthermore, our results show a practical use case of inconvenient interaction [55] (or uncomfortable interaction

[56,57]). Although most human-computer interaction studies focus on increasing efficiency, some previous studies addressed the need for and the value of inconvenient and uncomfortable interaction. Specifically, Rekimoto and Tsujita [55] suggested that inconvenient design can work in the domain of behavior change. For example, the refrigerator that does not open until the user smiles and the inconvenient microwave that requires the user to perform physical exercises during operation are examples of inconvenient interaction [55]. Unique usage contexts of wake-up tasking are closely related to the principles of inconvenient interaction. Even though the user likely feels inconvenienced by wake-up tasks, these tasks provide immediate and long-term benefits (eg, regular daily routine, positive mood, and physical health) that motivate the user to maintain the behavior.

### **Wake-Up Task Types**

This study discovered several differences between the 2 wake-up tasks. First, users using the *math\_task* tended to take a shorter time to dismiss an alarm than those using the *picture\_task*. Similar observations on alarm usage with 2 wake-up tasks were reported in an earlier study based on task-based alarm app usage logs [44]. However, the time between dismissing an alarm and performing the squats tended to be much shorter in the *picture\_task* group than in the *math\_task* group. Similarly, in the GEE analysis, the *picture\_task* group showed a marginally significant relationship with the elapsed time to start behavior, but the *math\_task* group did not. Finally, there were several differences in the workload demands of the 2 tasks. The total workload for the *picture\_task* group was slightly higher than that of the *math\_task* group. Furthermore, although the *picture\_task* group experienced more physical demand and frustration, the *math\_task* group experienced a more demanding mental workload.

These results likely indicate that tasks to elicit behavior change should be designed according to the direction of the anticipated change. Inserting another task between behaviors can be designed in both ways: preventing a particular behavior or preparing for the next. For example, some studies used interruption tasks for behavior change, such as typing digits to limit smartphone use [58], but wake-up tasking prepares the user for more productive morning behavior. The task function and mechanism can differ depending on the behavior goal. Furthermore, the task's workload amount and type can be essential for task-based behavior change. In this study, the target behavior was a form of physical activity, and participants likely had to move to a location where they could do the required squat exercises. Therefore, the target behavior would be more suitable for those completing a *picture\_task*, which mainly demands a physical workload and similarly moves the users to another place (as they need to capture a target picture). Further studies are necessary to identify the relationship between wake-up tasks and specific target behaviors for a more detailed understanding. It would also be extremely helpful if other types of wake-up tasks (eg, solving a puzzle) and morning behaviors (eg, reading an article) were studied further.

### **Factors Influencing Early Morning Behavior**

We also found that common elements for general behavior change (ie, TPB variables) were significantly related to the target behavior success in the morning. For example, *TPB\_intention* was closely significantly and positively associated with early success, and the increase in *TPB\_subjective\_norm* was significantly related to decreased behavior performance. However, the results address the need to consider the contextual characteristics of the morning hours because these general variables tended to contribute differently to morning contexts. For example, the magnitude of *TPB\_intention* is smaller than that of the other factors (ie, task usage and alarm usage). Moreover, the correlation direction of *TPB\_subjective\_norm* differed from those obtained in other contexts, such as recycling [59] or electronic learning [60]. Therefore, this study is in line with earlier studies that addressed the necessity of considering the contexts of the target behavior [61].

Sleep inertia characterizes the early morning hours. Moreover, people do not usually have much spare time in the early morning because their daily routine begins soon. Therefore, delayed waking up and starting of the target behavior can make users run out of available time and are likely to demotivate them. This study addresses three factors for successful morning behavior change: (1) waking up on time, (2) escaping from sleep inertia, and (3) quickly starting the desired target behavior.

First, it is essential to wake up on time to maintain the anticipated behavior in the early morning. The qualitative study revealed that when the target behavior failed, it was usually due to the user having overslept (see the comment by P7 for reference). Furthermore, most participants with wake-up tasks commented that the tasks helped make them wake up on time and begin the target behavior (see the comment by P15 and P34 for reference). Second, getting ready for the target morning behavior is necessary to escape sleep inertia. The GEE analysis results indicate that the performance of morning behavior decreased as the elapsed time to dismiss became longer, possibly due to oversleeping or a delay in waking up. Even though the participants woke up on time and dismissed their alarms, some remained half-asleep in bed and failed to do the desired target behavior (squats). Waking up late or sleep inertia can negatively affect the success of the target behavior. Finally, having a quick start is another key to performing the target behavior in the morning. The *non\_task* group users tended to dismiss their alarms quickly, but it took a long time to start the target behavior after dismissing the alarm. As a result, the total elapsed time to the target behavior was not too different from that of the other task groups (even marginally lower than the *picture\_task* group). It also had a significantly lower success rate than the other task groups.

Such factors influencing morning behavior can be further extended into a morning routine. Some previous studies for behavior change examined a chain of subsequent behaviors [62,63], and designing behavior can effectively induce other behavior changes [64,65]. Based on this study's findings, which focused on the beginning of the early morning, we anticipate that designing subsequent morning behaviors as a morning

routine may aid in completing productive morning hours, referred to as the miracle morning [1].

### Limitations

This study should be viewed critically because it was conducted at a single site in South Korea for 2 weeks.

This study also used a single target behavior type, squat exercises. In future research, other target behaviors can be studied. For example, reading, regarded as a cognitive activity, could be a good candidate for future morning behavior research. We also received responses from participants who were telecommuting due to COVID-19. However, this is not believed to have significantly influenced the study results because the experiment was conducted in situ by allowing all the participants to set and use alarms at their will.

### Conclusions

Although many people have a positive intention for productive morning behavior, contextual characteristics, such as sleep inertia, make it difficult. This study presents a mechanism for morning behavior change using wake-up tasks in a mobile alarm. Wake-up tasks provide winding-up and nudging effects to help escape from sleep inertia quickly and start the target behavior without subsequent delays. This study's quantitative and qualitative results confirmed the effectiveness of wake-up tasks in consistently completing the target behavior in the early morning. As a result, this wake-up task mechanism can contribute to extending the existing behavior change research to consider the morning hour context by shedding light on morning hour behaviors.

### Acknowledgments

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (Ministry of Science and ICT) (No. RS-2022-00155885, Artificial Intelligence Convergence Innovation Human Resources Development (Hanyang University ERICA)).

### Conflicts of Interest

None declared.

Multimedia Appendix 1

Results of the presurvey.

[[DOCX File , 23 KB - formative\\_v6i9e39497\\_app1.docx](#) ]

### References

1. Elrod H. *The Miracle Morning: The 6 Habits That Will Transform Your Life Before 8AM*. United Kingdom: John Murray Publishers; 2016.
2. Spall B, Xander M. *My Morning Routine: How Successful People Start Every Day Inspired*. United Kingdom: Penguin Books Ltd; 2018.
3. Kuismaa-Schildt M, Liukkonen J, Vuong M, Nyman K, Häkkinen K, Häkkinen A. Effects of morning vs. evening combined strength and endurance training on physical performance, sleep and well-being. *Chronobiol Int* 2019 Jun;36(6):811-825. [doi: [10.1080/07420528.2019.1592184](#)] [Medline: [30950283](#)]
4. Souissi H, Chtourou H, Chaouachi A, Dogui M, Chamari K, Souissi N, et al. The effect of training at a specific time-of-day on the diurnal variations of short-term exercise performances in 10- to 11-year-old boys. *Pediatr Exerc Sci* 2012 Feb;24(1):84-99. [doi: [10.1123/pes.24.1.84](#)] [Medline: [22433267](#)]
5. Souissi N, Gauthier A, Sesboüé B, Larue J, Davenne D. Effects of regular training at the same time of day on diurnal fluctuations in muscular performance. *J Sports Sci* 2002 Nov;20(11):929-937. [doi: [10.1080/026404102320761813](#)] [Medline: [12430993](#)]
6. Hobson RM, Clapp EL, Watson P, Maughan RJ. Exercise capacity in the heat is greater in the morning than in the evening in man. *Med Sci Sports Exerc* 2009 Jan;41(1):174-180. [doi: [10.1249/MSS.0b013e3181844e63](#)] [Medline: [19092692](#)]
7. Gomez AM, Gomez C, Aschner P, Veloza A, Muñoz O, Rubio C, et al. Effects of performing morning versus afternoon exercise on glycemic control and hypoglycemia frequency in type 1 diabetes patients on sensor-augmented insulin pump therapy. *J Diabetes Sci Technol* 2015 May;9(3):619-246 [FREE Full text] [doi: [10.1177/1932296814566233](#)] [Medline: [25555390](#)]
8. Sedliak M, Finni T, Cheng S, Kraemer WJ, Häkkinen K. Effect of time-of-day-specific strength training on serum hormone concentrations and isometric strength in men. *Chronobiol Int* 2007 Jul;24(6):1159-1177. [doi: [10.1080/07420520701800686](#)] [Medline: [18075805](#)]
9. Rothbard NP, Wilk SL. Waking up on the right or wrong side of the bed: start-of-workday mood, work events, employee affect, and performance. *Acad Manag J* 2011 Oct;54(5):959-980. [doi: [10.5465/amj.2007.0056](#)]
10. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychol Rev* 2015 Aug;9(3):323-344 [FREE Full text] [doi: [10.1080/17437199.2014.941722](#)] [Medline: [25104107](#)]

11. Thomas AR. *Intentional Behavior: An Approach to Human Motivation*. New York, NY: Ronald Press; 1970.
12. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *Am Psychol* 2002 Sep;57(9):705-717. [doi: [10.1037//0003-066x.57.9.705](https://doi.org/10.1037//0003-066x.57.9.705)] [Medline: [12237980](https://pubmed.ncbi.nlm.nih.gov/12237980/)]
13. Fogg BJ. A behavior model for persuasive design. In: *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*. 2009 Presented at: 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, CA p. 1-7. [doi: [10.1145/1541948.1541999](https://doi.org/10.1145/1541948.1541999)]
14. Fogg BJ. The behavior grid: 35 ways behavior can change. In: *Persuasive '09: Proceedings of the 4th International Conference on Persuasive Technology*. 2009 Presented at: 4th International Conference on Persuasive Technology; April 26-29, 2009; Claremont, CA p. 1-5. [doi: [10.1145/1541948.1542001](https://doi.org/10.1145/1541948.1542001)]
15. Prochaska JO, DiClemente CC. *The Transtheoretical Approach: Crossing Traditional Boundaries of Therapy*. Melbourne, FL: Krieger Pub Co; 1994.
16. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. *Am Psychol* 1992 Sep;47(9):1102-1114. [doi: [10.1037//0003-066x.47.9.1102](https://doi.org/10.1037//0003-066x.47.9.1102)] [Medline: [1329589](https://pubmed.ncbi.nlm.nih.gov/1329589/)]
17. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *J Med Internet Res* 2017 Jun;19(6):e232 [FREE Full text] [doi: [10.2196/jmir.7126](https://doi.org/10.2196/jmir.7126)] [Medline: [28663162](https://pubmed.ncbi.nlm.nih.gov/28663162/)]
18. Prestwich A, Webb TL, Conner M. Using theory to develop and test interventions to promote changes in health behaviour: evidence, issues, and recommendations. *Curr Opin Psychol* 2015 Oct;5:1-5. [doi: [10.1016/j.copsyc.2015.02.011](https://doi.org/10.1016/j.copsyc.2015.02.011)]
19. Fogg BJ. Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002 Dec;2002:5. [doi: [10.1145/764008.763957](https://doi.org/10.1145/764008.763957)]
20. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. *Am J Prev Med* 2016 Nov;51(5):825-832 [FREE Full text] [doi: [10.1016/j.amepre.2016.06.013](https://doi.org/10.1016/j.amepre.2016.06.013)] [Medline: [27745682](https://pubmed.ncbi.nlm.nih.gov/27745682/)]
21. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. *Commun Assoc Inf Syst* 2009 Mar;24:485-500. [doi: [10.17705/1CAIS.02428](https://doi.org/10.17705/1CAIS.02428)]
22. Oulasvirta A, Rattenbury T, Ma L, Raita E. Habits make smartphone use more pervasive. *Pers Ubiquit Comput* 2011 Jun;16(1):105-114. [doi: [10.1007/s00779-011-0412-2](https://doi.org/10.1007/s00779-011-0412-2)]
23. Consolvo S, Everitt K, Smith I, Landay JA. Design requirements for technologies that encourage physical activity. In: *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2006 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 22-26, 2006; Montreal, QC p. 457-466. [doi: [10.1145/1124772.1124840](https://doi.org/10.1145/1124772.1124840)]
24. Rabbi M, Li K, Yan HY, Hall K, Klasnja P, Murphy S. ReVibe: a context-assisted evening recall approach to improve self-report adherence. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2019 Dec;3(4):1-27 [FREE Full text] [doi: [10.1145/3369806](https://doi.org/10.1145/3369806)] [Medline: [34164595](https://pubmed.ncbi.nlm.nih.gov/34164595/)]
25. Chung C, Wang Q, Schroeder J, Cole A, Zia J, Fogarty J, et al. Identifying and planning for individualized change: patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2019 Mar;3(1):1-27 [FREE Full text] [doi: [10.1145/3314394](https://doi.org/10.1145/3314394)] [Medline: [31080941](https://pubmed.ncbi.nlm.nih.gov/31080941/)]
26. Costa J, Guimbretière F, Jung MF, Choudhury T. BoostMeUp: improving cognitive performance in the moment by unobtrusively regulating emotions with a smartwatch. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2019 Jun;3(2):1-23. [doi: [10.1145/3328911](https://doi.org/10.1145/3328911)]
27. Bruck D, Pisani DL. The effects of sleep inertia on decision-making performance. *J Sleep Res* 1999 Jun;8(2):95-103 [FREE Full text] [doi: [10.1046/j.1365-2869.1999.00150.x](https://doi.org/10.1046/j.1365-2869.1999.00150.x)] [Medline: [10389091](https://pubmed.ncbi.nlm.nih.gov/10389091/)]
28. Hilditch CJ, McHill AW. Sleep inertia: current insights. *Nat Sci Sleep* 2019;11:155-165 [FREE Full text] [doi: [10.2147/NSS.S188911](https://doi.org/10.2147/NSS.S188911)] [Medline: [31692489](https://pubmed.ncbi.nlm.nih.gov/31692489/)]
29. Tassi P, Muzet A. Sleep inertia. *Sleep Med Rev* 2000 Aug;4(4):341-353. [doi: [10.1053/smr.2000.0098](https://doi.org/10.1053/smr.2000.0098)] [Medline: [12531174](https://pubmed.ncbi.nlm.nih.gov/12531174/)]
30. Dawson D, Ferguson SA, Vincent GE. Safety implications of fatigue and sleep inertia for emergency services personnel. *Sleep Med Rev* 2021 Feb;55:101386. [doi: [10.1016/j.smr.2020.101386](https://doi.org/10.1016/j.smr.2020.101386)] [Medline: [33027747](https://pubmed.ncbi.nlm.nih.gov/33027747/)]
31. Trotti LM. Waking up is the hardest thing I do all day: sleep inertia and sleep drunkenness. *Sleep Med Rev* 2017 Oct;35:76-84 [FREE Full text] [doi: [10.1016/j.smr.2016.08.005](https://doi.org/10.1016/j.smr.2016.08.005)] [Medline: [27692973](https://pubmed.ncbi.nlm.nih.gov/27692973/)]
32. Dinges DF, Orne MT, Orne EC. Assessing performance upon abrupt awakening from naps during quasi-continuous operations. *Behav Res Methods Instrum Comput* 1985 Jan;17(1):37-45. [doi: [10.3758/bf03200895](https://doi.org/10.3758/bf03200895)]
33. Jewett ME, Wyatt JK, Ritz-De Cecco A, Khalsa SB, Dijk DJ, Czeisler CA. Time course of sleep inertia dissipation in human performance and alertness. *J Sleep Res* 1999 Mar;8(1):1-8 [FREE Full text] [doi: [10.1111/j.1365-2869.1999.00128.x](https://doi.org/10.1111/j.1365-2869.1999.00128.x)] [Medline: [10188130](https://pubmed.ncbi.nlm.nih.gov/10188130/)]
34. Sallinen M, Härmä M, Akerstedt T, Rosa R, Lillqvist O. Promoting alertness with a short nap during a night shift. *J Sleep Res* 1998 Dec;7(4):240-247 [FREE Full text] [doi: [10.1046/j.1365-2869.1998.00121.x](https://doi.org/10.1046/j.1365-2869.1998.00121.x)] [Medline: [9844850](https://pubmed.ncbi.nlm.nih.gov/9844850/)]
35. Tassi P, Nicolas A, Dewasmes G, Eschenlauer R, Ehrhart J, Salame P, et al. Effects of noise on sleep inertia as a function of circadian placement of a one-hour nap. *Percept Mot Skills* 1992 Aug;75(1):291-302. [doi: [10.2466/pms.1992.75.1.291](https://doi.org/10.2466/pms.1992.75.1.291)] [Medline: [1528684](https://pubmed.ncbi.nlm.nih.gov/1528684/)]

36. Hilditch CJ, Dorrian J, Banks S. Time to wake up: reactive countermeasures to sleep inertia. *Ind Health* 2016 Dec;54(6):528-541 [FREE Full text] [doi: [10.2486/indhealth.2015-0236](https://doi.org/10.2486/indhealth.2015-0236)] [Medline: [27193071](https://pubmed.ncbi.nlm.nih.gov/27193071/)]
37. McFarlane SJ, Garcia JE, Verhagen DS, Dyer AG. Alarm tones, music and their elements: analysis of reported waking sounds to counteract sleep inertia. *PLoS One* 2020 Jan;15(1):e0215788 [FREE Full text] [doi: [10.1371/journal.pone.0215788](https://doi.org/10.1371/journal.pone.0215788)] [Medline: [31990906](https://pubmed.ncbi.nlm.nih.gov/31990906/)]
38. McFarlane SJ, Garcia JE, Verhagen DS, Dyer AG. Auditory countermeasures for sleep inertia: exploring the effect of melody and rhythm in an ecological context. *Clocks Sleep* 2020 Jun;2(2):208-224 [FREE Full text] [doi: [10.3390/clockssleep2020017](https://doi.org/10.3390/clockssleep2020017)] [Medline: [33089201](https://pubmed.ncbi.nlm.nih.gov/33089201/)]
39. McFarlane SJ, Garcia JE, Verhagen DS, Dyer AG. Alarm tones, voice warnings, and musical treatments: a systematic review of auditory countermeasures for sleep inertia in abrupt and casual awakenings. *Clocks Sleep* 2020 Oct;2(4):416-433 [FREE Full text] [doi: [10.3390/clockssleep2040031](https://doi.org/10.3390/clockssleep2040031)] [Medline: [33118526](https://pubmed.ncbi.nlm.nih.gov/33118526/)]
40. Hilditch CJ, Wong LR, Bathurst NG, Feick NH, Pradhan S, Santamaria A, et al. Rise and shine: the use of polychromatic short-wavelength-enriched light to mitigate sleep inertia at night following awakening from slow-wave sleep. *J Sleep Res* 2022 Jan:e13558. [doi: [10.1111/jsr.13558](https://doi.org/10.1111/jsr.13558)] [Medline: [35102669](https://pubmed.ncbi.nlm.nih.gov/35102669/)]
41. Kaplan KA, Talavera DC, Harvey AG. Rise and shine: a treatment experiment testing a morning routine to decrease subjective sleep inertia in insomnia and bipolar disorder. *Behav Res Ther* 2018 Dec;111:106-112. [doi: [10.1016/j.brat.2018.10.009](https://doi.org/10.1016/j.brat.2018.10.009)] [Medline: [30399503](https://pubmed.ncbi.nlm.nih.gov/30399503/)]
42. Kovac K, Vincent GE, Paterson JL, Aisbett B, Reynolds AC, Ferguson SA. Can an increase in noradrenaline induced by brief exercise counteract sleep inertia? *Chronobiol Int* 2020 Sep;37(9-10):1474-1478. [doi: [10.1080/07420528.2020.1803900](https://doi.org/10.1080/07420528.2020.1803900)] [Medline: [32946288](https://pubmed.ncbi.nlm.nih.gov/32946288/)]
43. Oh KT, Shin J, Kim J, Ko M. Analysis of a wake-up task-based mobile alarm app. *Appl Sci* 2020 Jun;10(11):3993. [doi: [10.3390/app10113993](https://doi.org/10.3390/app10113993)]
44. Oh KT, Shin J, Kim J, Song IS, Lee U, Park E, et al. Wake-up task: understanding users in task-based mobile alarm app. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-6. [doi: [10.1145/3290607.3312765](https://doi.org/10.1145/3290607.3312765)]
45. Braun V, Clarke V. Thematic analysis. In: *APA Handbook of Research Methods in Psychology*. Washington, DC: American Psychological Association; 2012.
46. Bradski G. Dr. Dobb's Journal: Software Tools for the Professional Programmer. San Francisco, CA: CMP Media Inc; 2000.
47. Kayrouz R, Dear BF, Karin E, Fogliati VJ, Titov N. A pilot study of a clinician-guided internet-delivered cognitive behavioural therapy for anxiety and depression among Arabs in Australia, presented in both English and Arabic languages. *Internet Interv* 2016 Sep;5:5-11. [doi: [10.1016/j.invent.2016.06.002](https://doi.org/10.1016/j.invent.2016.06.002)] [Medline: [30135801](https://pubmed.ncbi.nlm.nih.gov/30135801/)]
48. Paik MC. Repeated measurement analysis for nonnormal data in small samples. *Commun Stat - Simul Comput* 1988 Jan;17(4):1155-1171. [doi: [10.1080/03610918808812718](https://doi.org/10.1080/03610918808812718)]
49. Teerenstra S, Lu B, Preisser JS, van Achterberg T, Borm GF. Sample size considerations for GEE analyses of three-level cluster randomized trials. *Biometrics* 2010 Dec;66(4):1230-1237 [FREE Full text] [doi: [10.1111/j.1541-0420.2009.01374.x](https://doi.org/10.1111/j.1541-0420.2009.01374.x)] [Medline: [20070297](https://pubmed.ncbi.nlm.nih.gov/20070297/)]
50. Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process* 1991 Dec;50(2):179-211. [doi: [10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)]
51. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv Psychol* 1988;52:139-183. [doi: [10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)]
52. Diggle P, Heagerty P, Liang KY, Zeger S. *Analysis of Longitudinal Data*. United Kingdom: Oxford university Press; 2002.
53. Amini P, Najafi-Vosough R, Ghaleiha A, Mahjub H. Evaluation of bipolar disorder in several relapses over time using generalized estimating equations. *Iran J Psychiatry* 2017 Jul;12(3):182-187 [FREE Full text] [Medline: [29062369](https://pubmed.ncbi.nlm.nih.gov/29062369/)]
54. Nam JH, Lim MS, Choi HK, Kim JY, Kim SK, Oh SS, et al. Factors increasing the risk for psychosocial stress among Korean adults living in rural areas: using generalized estimating equations and mixed models. *Ann Occup Environ Med* 2017 Oct;29(1):53 [FREE Full text] [doi: [10.1186/s40557-017-0209-5](https://doi.org/10.1186/s40557-017-0209-5)] [Medline: [29118991](https://pubmed.ncbi.nlm.nih.gov/29118991/)]
55. Rekimoto J, Tsujita H. Inconvenient interactions: an alternative interaction design approach to enrich our daily activities. In: *AVI '14: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. 2014 Presented at: International Working Conference on Advanced Visual Interfaces; May 27-29, 2014; Como, Italy p. 225-228. [doi: [10.1145/2598153.2598196](https://doi.org/10.1145/2598153.2598196)]
56. Benford S, Greenhalgh C, Giannachi G, Walker B, Marshall J, Rodden T. Uncomfortable interactions. In: *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, TX p. 2005-2014. [doi: [10.1145/2207676.2208347](https://doi.org/10.1145/2207676.2208347)]
57. Halbert H, Nathan LP. Designing for discomfort: Supporting critical reflection through interactive tools. In: *CSCW '15: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 2015 Presented at: 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; March 14-18, 2015; Vancouver, BC p. 349-360. [doi: [10.1145/2675133.2675162](https://doi.org/10.1145/2675133.2675162)]

58. Kim J, Park J, Lee H, Ko M, Lee U. LocknType: lockout task intervention for discouraging smartphone app use. In: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-12. [doi: [10.1145/3290605.3300927](https://doi.org/10.1145/3290605.3300927)]
59. Wan C, Shen GQ, Choi S. Experiential and instrumental attitudes: interaction effect of attitude and subjective norm on recycling intention. *J Environ Psychol* 2017 Jun;50:69-79. [doi: [10.1016/j.jenvp.2017.02.006](https://doi.org/10.1016/j.jenvp.2017.02.006)]
60. Cheung R, Vogel D. Predicting user acceptance of collaborative technologies: an extension of the technology acceptance model for e-learning. *Comput Educ* 2013 Apr;63:160-175. [doi: [10.1016/j.compedu.2012.12.003](https://doi.org/10.1016/j.compedu.2012.12.003)]
61. Ko M, Choi S, Yatani K, Lee U. Lock n' LoL: group-based limiting assistance app to mitigate smartphone distractions in group activities. In: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 998-1010. [doi: [10.1145/2858036.2858568](https://doi.org/10.1145/2858036.2858568)]
62. Iqbal ST, Horvitz E. Disruption and recovery of computing tasks: field study, analysis, and directions. In: CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 29-May 3, 2007; San Jose, CA p. 677-686. [doi: [10.1145/1240624.1240730](https://doi.org/10.1145/1240624.1240730)]
63. Lee U, Lee J, Ko M, Lee C, Kim Y, Yang S, et al. Hooked on smartphones: an exploratory study on smartphone overuse among college students. In: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: 2014 CHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, ON p. 2327-2336. [doi: [10.1145/2556288.2557366](https://doi.org/10.1145/2556288.2557366)]
64. Hiniker A, Heung SS, Hong SR, Kientz JA. Coco's videos: An empirical investigation of video-player design features and children's media use. In: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC p. 1-13. [doi: [10.1145/3173574.3173828](https://doi.org/10.1145/3173574.3173828)]
65. Hiniker A, Hong SR, Kohno T, Kientz JA. MyTime: designing and evaluating an intervention for smartphone non-use. In: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 4746-4757. [doi: [10.1145/2858036.2858403](https://doi.org/10.1145/2858036.2858403)]

## Abbreviations

**GEE:** generalized estimating equation

**MAE:** mean absolute error

**MAP:** motivation, ability, and prompt

**NASA-TLX:** NASA-Task Load Index

**TPB:** theory of planned behavior

*Edited by A Mavragani; submitted 25.05.22; peer-reviewed by Q Chen; comments to author 16.06.22; revised version received 11.08.22; accepted 29.08.22; published 21.09.22.*

*Please cite as:*

*Oh KT, Ko J, Shin J, Ko M*

*Using Wake-Up Tasks for Morning Behavior Change: Development and Usability Study*

*JMIR Form Res* 2022;6(9):e39497

URL: <https://formative.jmir.org/2022/9/e39497>

doi: [10.2196/39497](https://doi.org/10.2196/39497)

PMID: [36129742](https://pubmed.ncbi.nlm.nih.gov/36129742/)

©Kye Taek Oh, Jisu Ko, Jaemyung Shin, Minsam Ko. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 21.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Tough Talks Virtual Simulation HIV Disclosure Intervention for Young Men Who Have Sex With Men: Development and Usability Testing

Lisa B Hightow-Weidman<sup>1</sup>, MD, MPH; Kathryn Muessig<sup>1</sup>, PhD; Zach Soberano<sup>1</sup>, BA; Matthew T Rosso<sup>1</sup>, MPH; Andrew Currie<sup>2</sup>, BS; Margo Adams Larsen<sup>2</sup>, PhD; Kelly Knudtson<sup>3</sup>, MPH; Alyssa Vecchio<sup>4</sup>, MD

<sup>1</sup>Institute for Global Health and Infectious Diseases, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

<sup>2</sup>Virtually Better Inc, Decatur, GA, United States

<sup>3</sup>School of Medicine, University of Washington, Seattle, WA, United States

<sup>4</sup>Department of Family & Community Medicine, University of New Mexico, Albuquerque, NM, United States

**Corresponding Author:**

Lisa B Hightow-Weidman, MD, MPH  
Institute for Global Health and Infectious Diseases  
University of North Carolina at Chapel Hill  
130 Mason Farm Rd  
Chapel Hill, NC, 27599  
United States  
Phone: 1 919 843 0033  
Email: [lisa\\_hightow@med.unc.edu](mailto:lisa_hightow@med.unc.edu)

## Abstract

**Background:** HIV status disclosure is an important decision with barriers specific to young men who have sex with men (YMSM), who have the highest rates of new HIV infections in the United States. Behavioral and social determinants of the difficulty to disclose can include fear of rejection, stigma, loss of financial stability, and lack of communication skills. Once able to disclose, a person may have increased access to social support and improved informed risk reduction conversations and medication adherence. Despite the known challenges and advantages of disclosure, there are few effective tools supporting this behavior.

**Objective:** To address this gap in disclosure interventions, the Tough Talks (TT) app, an mHealth intervention using artificial intelligence (AI)-facilitated role-playing scenarios, was developed for YMSM. This paper reports stages of development of the integrated app and results of the usability testing.

**Methods:** Building on the successful development and testing of a stand-alone interactive dialogue feature in phases 1-3, we conducted additional formative research to further refine and enhance the disclosure scenarios and develop and situate them within the context of a comprehensive intervention app to support disclosure. We assessed the new iteration for acceptability and relevance in a usability study with 8 YMSM with HIV. Participants completed a presurvey, app modules, and a semistructured qualitative interview.

**Results:** TT content and activities were based on social cognitive theory and disclosure process model framework and expanded to a 4-module curriculum. The AI-facilitated scenarios used dialogue from an utterance database developed using language crowdsourced through a comic book contest. In usability testing, YMSM reported high satisfaction with TT, with 98% (31/33) of activities receiving positive ratings. Participants found the AI-facilitated scenarios and activities to be representative and relevant to their lived experiences, although they noted difficulty having nuanced disclosure conversations with the AI.

**Conclusions:** TT was an engaging and practical intervention for self-disclosure among YMSM with HIV. Facilitating informed disclosure decisions has the potential to impact engagement in sexual risk behaviors and HIV care. More information is needed about the ideal environment, technical assistance, and clinical support for an mHealth disclosure intervention. TT is being tested as a scalable intervention in a multisite randomized controlled trial to address outstanding questions on accessibility and effect on viral suppression.

**Trial Registration:** ClinicalTrials.gov NCT03414372; <https://clinicaltrials.gov/ct2/show/NCT03414372>

**KEYWORDS**

HIV; virtual reality; status disclosure; prevention; young men who have sex with men; artificial intelligence; medication adherence; transmission; viral load; men; sex; development; usability; testing; virtual simulation; young men; United States; behavior; social determinants

## *Introduction*

HIV status disclosure is an important and often challenging personal decision for youth with HIV. Self-disclosure, or the sharing of personal information such as one's HIV status, with others can be an integral part of social interaction [1]. Barriers to disclosure for youth with HIV include fear of rejection, stigma, and loss of financial stability [2,3]. Moreover, youth with HIV may lack the skills needed to effectively communicate with others regarding their status [2,3]. Studies conducted with men who have sex with men (MSM) in the United States have shown that 30% to 40% of persons living with HIV inconsistently disclose their status to sex partners, with higher odds of condomless anal intercourse (CAI) and increased number of partners among those who do not disclose [4-6]. Compelling disclosure-support interventions for youth with HIV should be grounded in their experiences of disclosure and address the barriers they face.

Status disclosure among youth with HIV is influenced by multiple factors, including the relationship to the person to whom they disclose, length of time since receiving an HIV diagnosis, and means of HIV acquisition (eg, behavioral vs perinatal transmission) [2,3]. Youth with HIV who disclose are more likely to engage in safer sex practices and report better social support [7,8]. Further, disclosure has been shown to impact subsequent mental health benefits and may improve engagement in HIV care and medication adherence [2,9]. Given the low rate of antiretroviral therapy (ART) adherence among youth with HIV contrasted with the benefits of consistent viral suppression including reduced rate of transmission to sex partners, HIV status disclosure may play an integral role in ongoing efforts to end the HIV epidemic [10-12].

With new HIV infections in the United States disproportionately affecting young MSM (YMSM), it is paramount that efforts to improve disclosure self-efficacy prioritize this at-risk group [13-15]. Despite the public health and personal benefits of disclosure, there is currently a lack of effective interventions; a 2019 systematic review identified 1455 publications with disclosure interventions, 8 that met the inclusion criteria, of which only 5 were efficacious in promoting disclosure to sex partners [16].

Tough Talks (TT), a mobile health (mHealth) app-based intervention using artificial intelligence (AI)-facilitated role-playing scenarios for YMSM with HIV, was developed in 2019 to meet this need. The content developed for TT is based on social cognitive theory (SCT), and the formative research for the initial prototype has been previously described [17,18]. In summary, the formative research consisted of the first 3 phases of development that produced and tested 3 AI-facilitated

role-playing scenarios for YMSM to practice HIV status disclosure [17]. Pilot testing with 11 youth with HIV showed high acceptability for the scenarios; however, user feedback indicated the need for additional disclosure-focused content to support the AI-facilitated role-plays [17].

Here we describe the final phases of development and testing to create the fully developed, multicomponent TT. This includes the participant-recommended additional content, namely the SCT-based modular curriculum and crowdsourced expansion of disclosure scenarios, along with a paradata (ie, user intervention analytics) [18,19] reporting system that tracks user activities completed, number of log-ins, time spent, etc, and an administrator portal to move TT toward final clinical testing.

## *Methods*

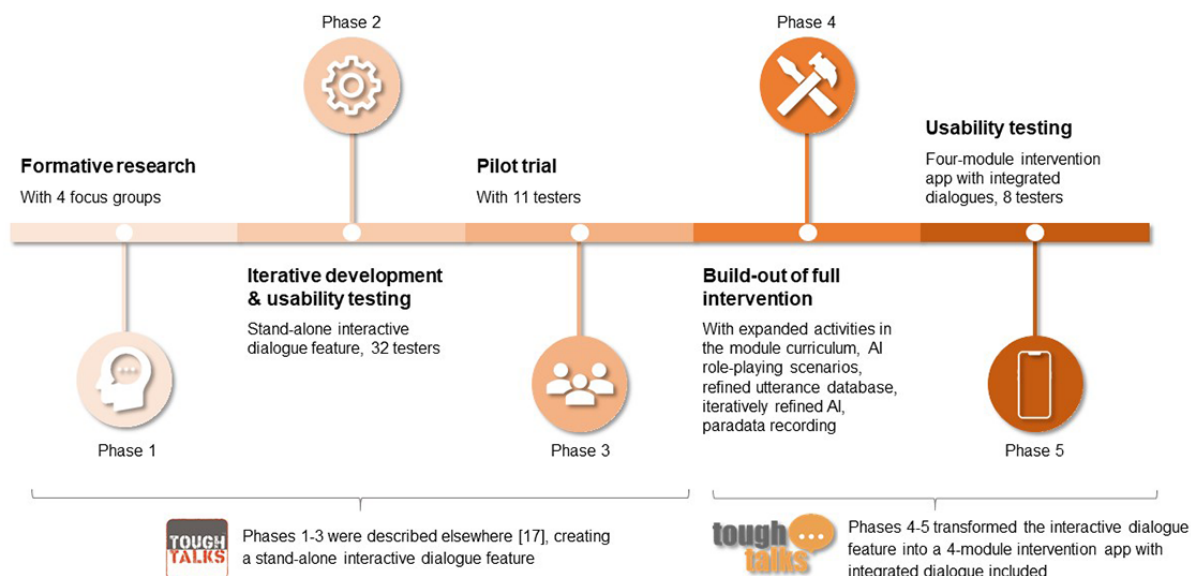
### **Overall**

TT was created through a collaboration between the Behavior and Technology Lab (BATLab) at the University of North Carolina at Chapel Hill (UNC-CH) and Virtually Better, Inc. (VBI). The interdisciplinary group drew upon expertise in a spectrum of fields from HIV prevention and clinical psychology to computer science and product development. Each stage of the formative work included feedback from YMSM to ensure the content, context, and function of the app was relevant and acceptable.

The process and results of designing this integrated, theory-based intervention are described by phases (Figure 1). Since the outcomes of each phase influence the development and methods of subsequent phases, an integrated description is presented. A stand-alone interactive dialogue feature was created and refined in phases 1 to 3 as previously presented [17]. Formative research in phase 1 began with participants from the 4 focus groups (n=58) describing past disclosure experiences, discussing disclosure strategies (barriers and facilitators) and working in pairs to develop real-life disclosure scenario conversations, and assisting in creating a stand-alone interactive dialogue feature. In phase 2, the dialogue feature was further refined through 3 rounds of usability testing with YMSM with HIV (n=32), and in phase 3, the feature was tested for preliminary efficacy with 11 new YMSM with HIV. The TT prototype at the end of phase 3 included an SCT-informed virtual dialogue feature to guide YMSM through HIV status disclosure conversations. This paper describes methods to supplement the dialogue feature with activities and educational content in response to participant feedback from earlier phases and align more closely with the SCT framework [20]. Specifically, development of the fully integrated app-based intervention and usability testing are described in phases 4 and 5, respectively.



**Figure 1.** Tough Talks timeline: development of Tough Talks app, phases 1-5.



### Phase 4 Activities

Additional AI-facilitated role-playing scenarios were developed to include more complex disclosure conversations and allow users to practice disclosing to a simulated sex partner. Authentic and realistic disclosure dialogues for these activities were sought using online crowdsourcing in the form of a comic book contest [21]. The online submission platform included 8 comic book panels that included dialogue prompts to initiate a disclosure dialogue. Panels with no introductory text were also included to encourage original submissions. Text from the situations in these panels are included in Table 1. The online comic book contest was conducted between November 2017 to February

2018 and advertised on social media (eg, Facebook, Grindr). The website provided specific submission instructions, including entry eligibility (eg, must include intelligible dialogue) and a brief tutorial on the submission process. Multiple entries were allowed and encouraged. Eligible participants were aged 16 to 29 years, born male and identify as male, and having had anal sex with another man or intended to in last 12 months. Submissions were judged by the BATLab study team and members of the youth advisory board on a scale of 1 to 5 (5 being the best) on the content of the conversations within the following 3 domains: relatability, uniqueness, and relevancy of the dialogue for the final TT disclosure scenarios.

**Table 1.** Online comic book contest disclosure scenario prompts.

Scenario	Introduction text to the disclosure situation
WTF <sup>a</sup> text	You’ve hooked up with this guy a couple times, usually when you see him out. It’s very casual. You receive a text from him. “Hey. My friend just told me you’re poz. WTF. Why didn’t you tell me???”
Catching feelings	You’re meeting up with a partner you have been dating for a few weeks. You two have not yet had sex, but you both want to. You really like him but are nervous that he won’t want to continue with you if he knows you are HIV positive.
Snooping	You’re in bed cuddling with a guy you just hooked up with. He gets up to go the bathroom and walks out holding a bottle of your HIV meds.
FWB <sup>b</sup> to Bae?	You’ve been hooking up with a partner randomly for about 2 months. You both start developing feelings for each other and hanging out more. You haven’t talked about each other’s status but you want to.
Protected?	You meet up with a partner you have been dating for a few weeks. You had protected sex with them a few nights ago. You two have not talked about status.
Club hook-up	You’re making out with a guy in the club and he pulls you toward the bathroom.
But I don’t like condoms	You’re about to hook up with a partner you just met after going back to his place. You realize you don’t have a condom and ask him for one. He says, “I don’t like condoms.”
Slide into those DMs <sup>c</sup>	You met a guy on Instagram and have been texting for about a month. It’s always really flirty, and you’ve started to really like him. You guys haven’t talked about HIV status.

<sup>a</sup>WTF: What the fuck.

<sup>b</sup>FWB: friends with benefits.

<sup>c</sup>DM: Direct Message.

## Ethics Approval

This study was approved (14-0345) by the institutional review board at UNC-CH.

## Results

---

### Overall

Module content and activities (Table 2) are grounded in the constructs of SCT and informed by the results of phases 1 to 3 [17]. Content addresses internalized HIV stigma, fear of disclosing one's status, fear of rejection, feelings of being overwhelmed, pressure to educate others about HIV, and past negative experiences with disclosure. Cognitive factors, specifically knowledge, expectations, and attitudes surrounding disclosure, are explored through videos, educational activities, and quizzes. We also added disclosure legal requirements by

state, easy to understand information about HIV, tips for how to talk to others about HIV, and safer sex facts. Environmental factors (ie, intersectional stigma, regional context, social norms) are addressed through play-through choose-your-own-adventure style narratives, and activities focused on disclosure context and timing. Behavioral factors to increase disclosure skills and self-efficacy are addressed with activities that include decisional balances around disclosure costs and benefits, sexual communication tips, and coping mechanisms for disclosing including through emotional regulation. These activities aligned with formative data in which participants described weighing the costs and benefits of disclosing their HIV status to intimate partners, friends, and family. For some, the social support made it worth it, while others learned how to accept the range the outcomes including those from people who were not as open or informed.

**Table 2.** Social cognitive theory–informed activities included in modules.

	Cognitive factors			Environmental factors			Behavioral factors		
	Knowledge	Expectations	Attitudes	Social norms	Access in community	Influence on others	Skills	Practice	Self-efficacy
0.1. Introduction and goals		✓							✓
1.0. CYOA <sup>a</sup> : It's like coming out...again		✓	✓	✓				✓	✓
1.1. Disclosure and state laws	✓			✓					
1.2. What is disclosure?	✓	✓							
1.3. Who needs to know?	✓	✓							
1.4. I am (blank)		✓	✓						
1.5. What would you do?		✓	✓	✓		✓	✓	✓	
1.6. Virtual disclosure practice		✓	✓		✓		✓	✓	✓
2.0. CYOA: The dating game...with a twist		✓	✓		✓			✓	✓
2.1. Your past experiences		✓	✓						
2.2. To disclose or not to disclose?		✓	✓			✓	✓	✓	
2.3. Is now the time to do it?	✓	✓		✓					
2.4. Right time, right place	✓	✓	✓			✓	✓	✓	
3.0. CYOA: He likes me...he likes me not		✓	✓	✓				✓	✓
3.1. Breaking the ice (subtly)	✓	✓	✓			✓	✓	✓	
3.2. Tell it over text	✓	✓	✓	✓		✓	✓		
3.3. Conversation starters	✓	✓	✓				✓	✓	
3.4. The cat's out of the bag	✓	✓	✓				✓	✓	
4.0. CYOA: How did you find out?		✓	✓	✓				✓	✓
4.1. Q&A <sup>b</sup> : HIV edition	✓	✓					✓	✓	
4.2. What are you willing to answer?	✓	✓	✓			✓	✓	✓	✓
4.3. How would you answer?	✓	✓	✓	✓			✓	✓	
4.4. What am I most afraid of?	✓	✓	✓	✓					
4.5. Get out while you can	✓	✓	✓			✓	✓	✓	
4.7. Reflections 2		✓	✓						✓

<sup>a</sup>CYOA: choose your own adventure.

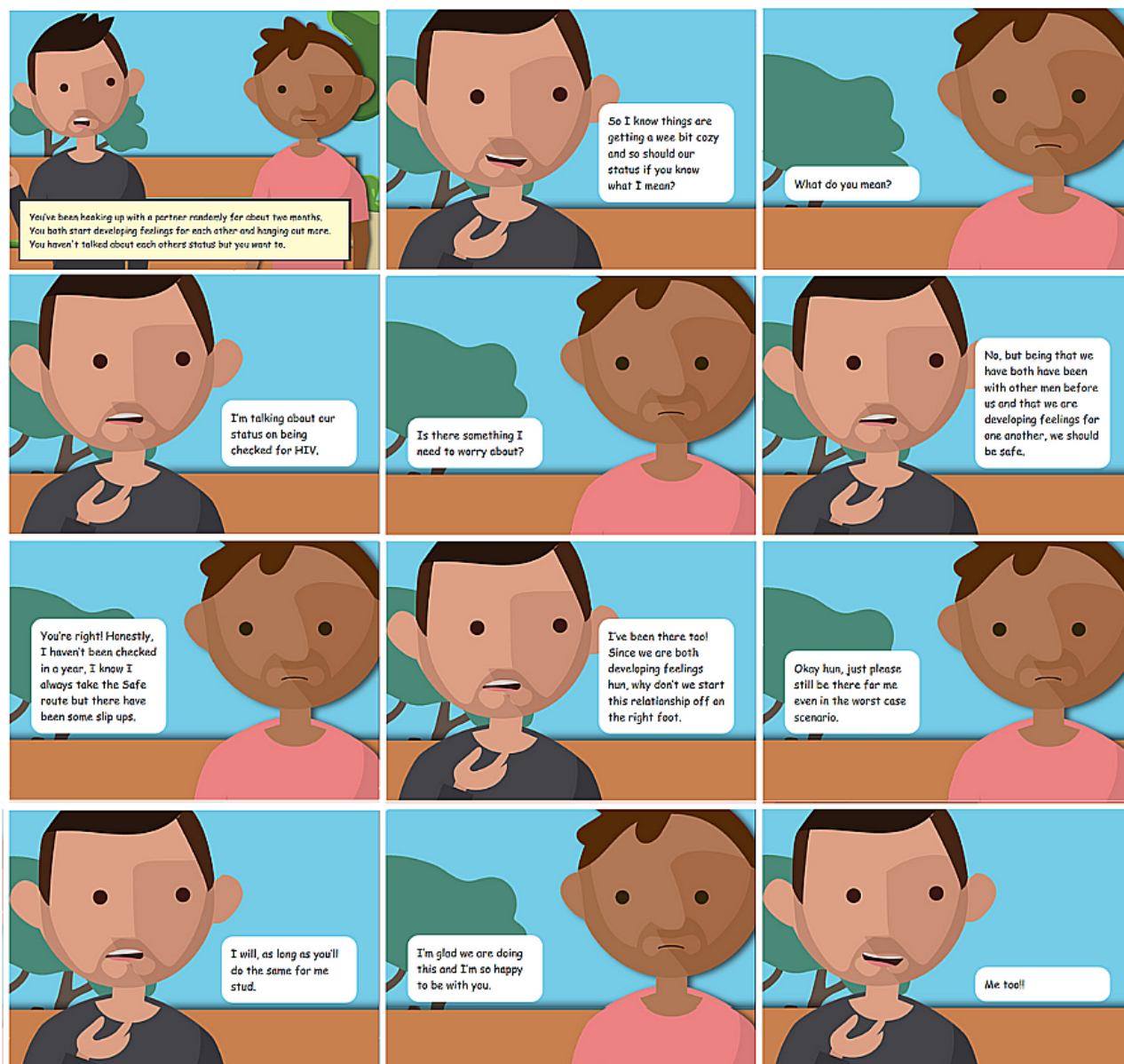
<sup>b</sup>Q&A: questions and answers.

### Expanded AI Scenarios: Comic Book Contest and Refinement of Utterance Database

A total of 76 eligible and consented participants created accounts on the contest website; 21 completed comics were submitted during the contest, including 16 daily winners, 4 weekly winners, and 1 grand prize winner (Figure 2). We incorporated the comics' relevant context, situations, and language into the module activities (eg, "Tell it over text") and the AI-facilitated role-playing disclosure scenarios. This created a richer selection

of responses for the utterance database ("Honestly, I haven't been checked in a year. I know I always take the safe route, but there have been some slip ups" as a response to a TT user's question about HIV status or testing), the collection of responses the AI system could use as the virtual character based on user conversation. The responses identified as reactions to disclosure were classified as being positive, negative, or neutral so the virtual character (ie, avatar) could be programmed and properly respond with a full range of reactions to the app user's status disclosure.

Figure 2. Online comic book contest winning disclosure dialogue.



### Development of an Integrated App-Based Intervention

The final intervention consists of 4 modules containing 24 activities and 8 AI-facilitated role-playing scenarios. Each module builds on the previous one in terms of disclosure knowledge, decision-making, and practice. All modules (Understanding disclosure, Should I disclose?, How do I disclose?, Preparing for the outcome) begin with a choose-your-own-adventure style game featuring a young man living with HIV navigating how to disclose to family, friends,

and romantic and sexual partners (Table 1). Modules also include 4 to 5 informational/skill-building activities, goal setting, and reflection activities. Modules also include an embedded AI-facilitated role-playing scenario with 2 variable partner reaction outcomes (positive, neutral, negative) (Table 3). Positive reactions were those where the avatar was explicitly supportive, neutral reactions were those in which the avatar was neither explicitly supportive nor unsupportive, and negative reactions were those in which the avatar was explicitly unsupportive.

**Table 3.** Artificial intelligence–facilitated role-play scenarios and partner reaction outcomes.

Module	Scenario	Have they previously had sex?	First reaction outcome	Second reaction outcome
1	You met a guy on Instagram and have been texting for about a month. It's always really flirty, and you've started to really like him. You guys haven't talked about HIV status.	No	Positive	Neutral
2	You're about to hook up with a partner you just met after going back to his place. You realize you don't have a condom and ask him for one. He says, "I don't like condoms."	No	Neutral	Negative
3	You meet up with a partner that you have been dating for a few weeks. You had protected anal sex with him using condoms a few nights ago. You two have not talked about status.	Yes	Negative	Positive
4	You've been hooking up with a partner randomly for about 2 months. You have had condomless sex. You both start developing feelings for each other and hanging out more. You haven't talked about each other's status, but you want to.	Yes, condomless	Positive (on PrEP <sup>a</sup> )	Neutral

<sup>a</sup>PrEP: preexposure prophylaxis.

### App Backend Database: Enhancing the Avatar Responses and Capturing Intervention Paradata

The SCT-based activities and disclosure scenarios collected from the comic book contest were integrated into TT by VBI. Members of the research team completed internal usability testing of TT, with particular attention to coding the AI to properly select an appropriate avatar response based on the conversation's tone and context. In the fully automated system, the natural language processing system [17] chooses a response it determines fits best for the simulated sex partner based on the scenario and TT user input. To address limitations to the AI program we identified such as its ability to understand complex sentences from the TT user, we developed and tested a semiautomated, "wizard-of-oz"–driven system in which the AI program provided suggested responses to a staff member through the administrator portal but allowed the staff member to either accept this suggestion or use another one to respond as the simulated character to the TT user. Of note, all the possible responses in the semiautomated version were the same as those contained in the automated version (eg, the staff member could not craft their own response). VBI created a backend database to track intervention paradata consisting of all participant activity conducted within the app (eg, length of time app was used, modules and activities completed, disclosure scenario chat logs). For the study team, VBI created an administrator portal for TT where the research administrator could view the coded user ID, date/time of last activity, and the cumulative number of activities completed. This backend feature captures the intervention paradata and allows research administrators to monitor TT use.

### Usability Testing of Fully Developed Intervention

Once the prototype was fully developed, a usability assessment of TT was conducted in North Carolina from July to October 2018. Eligibility criteria included being born male and identifying as male, being aged 16 to 29 years, self-reporting living with HIV, and being fluent in English, and having had at least one episode of CAI with another man or transgender woman in the past 6 months. Recruitment was primarily through flyers and referrals from local HIV clinics.

Study team members conducted usability sessions in private rooms at community- and clinic-based study sites. Each session lasted 60 to 90 minutes. Participants completed a presurvey (Multimedia Appendix 1), reviewed TT modules, and participated in semistructured qualitative interviews. The presurvey included questions on demographics, disclosure experience (ever disclosed to a sexual partner, anal or vaginal sex without disclosure in past 12 months), use of technology, and experience with AI-facilitated games/interventions ("Do you have any AI-related experience? This can be through online virtual worlds, games, etc."). Participants each reviewed 2 of the 4 modules. Participants were given a workbook that included 2 questions about every activity (liked activity and activity length) as well as a notes section for every activity.

Following the administration of the modules, participants completed a semistructured interview to rate TT on visual appeal, acceptability, ease of use, and user experience. In these interviews, participants were asked to assess the relevance and accuracy of the conversations as compared to their prior disclosure experiences (or anticipated experiences), emotional impact of using the program, how and where they could see themselves using the program, perceived usefulness of the program, and suggestions for improvements. Usability interview sessions were audiorecorded. Participants received \$50 for participation. Results of the survey and quantitative workbook responses were reviewed and analyzed in Excel (Microsoft Corp) for demographics and responses held by a majority of the participants. Thematic analysis of usability qualitative interview was conducted through multiple phases of reading and coding by study team members experienced in public health qualitative data analysis. Given the size of the cohort, qualitative data analysis software was not used for coding; we opted instead for a collaborative process of extracting the themes from interviews and workbooks.

### Participant Characteristics

A total of 8 YMSM living with HIV were enrolled (Table 4); mean age was 27.6 years, and all identified as Black. Mean time since HIV diagnosis was 4.6 years (range 5 months to 12 years). All participants were in-care, virally suppressed, and self-reported ≥90% ART adherence in the last month. All

reported anal sex in the last 3 months, 88% (7/8) of the YMSM reported CAI, and half (4/8, 50%) reported an STI in the last 6 months. Of the YMSM, 62% (6/8) believed people are rejected when they disclose; 88% (7/8) reported disclosing their status

to a sex partner. Among those reporting CAI, 57% (4/8) of the YMSM disclosed to some partners, 29% (2/8) disclosed to all partners, and 14% (1/8) did not disclose to partners.

**Table 4.** Participant characteristics (n=8).

	Value
Age (years), mean	27.6
<b>Race/ethnicity, n (%)</b>	
Black or African American, non-Hispanic	8 (100)
<b>Time since HIV diagnosis (years)</b>	
Average	4.6
Minimum	0.4
Maximum	12
Health care, n (%)	8 (100)
Virally suppressed, n (%)	8 (100)
Self-reported ART <sup>a</sup> adherence $\geq 90\%$ , n (%)	8 (100)
<b>Anal intercourse, n (%)</b>	
Past 3 months	8 (100)
Condomless anal intercourse	7 (88)
STI <sup>b</sup> in the last 6 months	4 (50)
Rejected/stigma with disclosure, n (%)	5 (63)
Disclosure to sex partner, n (%)	7 (88)
<b>Disclosure among CAI<sup>c</sup> (n=7), n (%)</b>	
Disclosed to some partners	4 (57)
Disclosed to all partners	2 (29)
Did not disclose to partners	1 (14)

<sup>a</sup>ART: antiretroviral therapy.

<sup>b</sup>STI: sexually transmitted infection.

<sup>c</sup>CAI: condomless anal intercourse.

## Usability Feedback

All participants reported high satisfaction with TT, with 93% (31/33) of activities rated positively. There was general ease with using TT and learning new information. The most common feedback on individual activities centered on activity length, specifically that the activities were too long. Participants reported that information provided in TT was useful and new to them. They requested more activities that incorporated information on disclosing to family, dating, and additional information pertaining to HIV and state disclosure laws.

Participants offered feedback on specific components of the intervention. They particularly enjoyed the choose-your-own-adventure style games since they related to the character's journey. Many mentioned that they had similar situations with their siblings or could see themselves having those conversations with friends. Many noted that they would repeat those activities if they had the chance. There were positive comments regarding the overall design and graphics.

Participants had mixed reactions and opinions about the AI-facilitated role-playing scenarios. While most participants found them to be the most interesting part of the app, specifically citing that they were "useful and realistic," 3 YMSM reported that receiving negative responses from the avatar could be emotionally difficult at times. One participant specifically commented that the intervention should provide more emotional support after the simulated sex partner reacts negatively to their disclosure. Two participants noted the AI-facilitated role-playing scenarios had noticeable bugs (eg, slowness or confusing simulated sex partner responses) but still found the activity useful.

The usability study participants had to type out the conversation with the simulated sex partner as opposed to being able to speak out loud. Many participants noted that this, among other factors, took away from the realistic nature of the disclosure conversation practice.

One participant appreciated the overall tone of TT and how it focused on giving back their identity and "reinforces being

yourself and that HIV doesn't affect you." Many participants noted that they would have found TT useful when they were first diagnosed instead of having to build their own skills to learn to disclose their status. They explained the challenges of disclosing to family, peers, and partners. One participant described the constant stress, even if at a low level, of not disclosing. Some of the fears of disclosure included partner rejection, family judgement, and physical violence.

### **Implications of the Usability Study to Further Intervention Design**

Changes to the design and content of the app based on participant feedback were implemented before launching the randomized controlled trial (RCT).

Specific changes are noted in [Table 5](#). We updated the titles and order of the activities and modules to make the sections feel less school-like or childish. Features, including buttons for forgotten passwords and returning to the home screen, were

modified to make TT more user-friendly. Many activities were reduced in length (eg, from around 5 minutes to 2-3 minutes) based on participant feedback. The final time for each module was reduced from 60 minutes to 45 minutes, and the overall TT intervention was shortened from 4 hours to 3 hours.

The most significant modifications proposed were to the AI technology. Speech recognition was added freeing participants from having to type responses. A chat log history box was added to the bottom of the screen to allow the user to review the disclosure discussion ([Figure 3](#)). Changes were made to address emotional concerns: (1) coach dialogue was enhanced to help guide users through the scenario; (2) reflection activities were added to postdisclosure scenarios to help users process information received during the scenario, and (3) access to virtual and in-person support as needed, including local and online resources for sexual and mental health and staff and clinic contact information, was confirmed.

**Table 5.** Changes to activities based on user feedback.

Module and activity	Title	Activity	Integrated feedback
<b>Introduction</b>			
0	Your coach	Meet the coach	
0.1	Welcome!	Coach explains what to expect in app	
0.2	Let's set some goals!	Set both short-term and long-term goals	Split goals into separate sections
<b>Module 1: Understanding disclosure</b>			
1	It's like coming out... again	CYOA <sup>a</sup> deciding if/how to disclose to a sister	Added animation to text boxes
1.1	What is disclosure?	Animated video explaining disclosure	Split into separate shorter videos
1.2	Disclosure and state laws	Differences in state laws about disclosure	Updated to include all 50 states
1.3	Who needs to know?	Quiz about who people need to disclose to	
1.4	I am (blank)...	Fill-in-the-blank about role of HIV stigma	Updated to be more age appropriate
1.5	What would you do?	Give advice to a friend about disclosure	Improved texting simulation
1.6	Virtual disclosure practice	Role-play status disclosure to a virtual partner	
<b>Module 2: Should I disclose?</b>			
2	The dating game...with a twist	CYOA about disclosing status on dating apps	Added animation to text boxes
2.1	Your past experiences	Reflect on previous disclosure experiences	Added colors to improve user experience
2.2	To disclose or not disclose?	Sort positive and negative outcomes of disclosure	Shortened activity and clarified language
2.3	Is now the time to do it?	Animated video about disclosure circumstances	Video added from previous section
2.4	Right time, right place	Pros and cons of different settings to disclose	
2.5	Virtual disclosure practice	Role-play status disclosure to a virtual partner	
2.6	Reflection	First impressions and revisit initial goals	
<b>Module 3: How do I disclose?</b>			
3	He likes me...he likes me not	CYOA about dating a coworker	Added animation to text boxes
3.1	Breaking the ice subtly	Animated video with tips for subtle disclosure	
3.2	Tell it over text	Examples of disclosure conversations via text	Improved texting simulation
3.3	Conversation starters—bracket challenge	Choose favorite disclosure conversation starters	Improved visual user experience
3.4	The cat's out of the bag	Consider how to respond to person finding out status	Added text entry box to be interactive
3.5	Virtual disclosure practice	Role-play status disclosure to a virtual partner	
<b>Module 4: Preparing for the outcome</b>			
4	How did you find out?	CYOA when a friend inadvertently finds out	Added animation to text boxes
4.1	Q&A <sup>b</sup> : HIV edition	Answers to FAQs <sup>c</sup> about HIV postdisclosure	Improved visual user experience
4.2	What are you willing to answer?	Self-reflection on what you'll personally answer	Added "other" text entry option
4.3	How would you answer?	Fill-in-the-blank comic-style conversation activity	
4.4	What am I most afraid of?	Recognize fears surrounding disclosure	
4.5	Get out while you can	Tips if disclosure goes badly	
4.6	Virtual disclosure practice	Role-play status disclosure to a virtual partner	
4.7	Reflection	Final check-in and revisit goals	

<sup>a</sup>CYOA: choose your own adventure.<sup>b</sup>Q&A: questions and answers<sup>c</sup>FAQ: frequently asked questions.



**Figure 3.** Screen shots of artificial intelligence–facilitated role plays.



## Discussion

### Principal Findings

TT, a behavioral intervention to practice HIV status disclosure, was developed over multiple phases of research and participant testing. Building on the successful development and testing of a virtual simulation prototype that demonstrated AI-driven scenarios were both feasible and acceptable to YMSM to practice HIV status disclosure with sex partners (phases 1 to 3) [17], we transformed the virtual prototype into a modular disclosure curriculum contained within an interventional mobile health app tested with 8 YMSM living with HIV (phases 4 and 5). Throughout the development process, the perspectives of YMSM were used to create a relevant and engaging intervention. Usability testing found that the intervention was both feasible and acceptable among members of the priority population, who reported learning new information after engaging in the TT activities.

Participants in the early phases of the intervention development confirmed the barriers and potential benefits of HIV status disclosure that have been discussed in previous literature [22,23]. However, they also identified personal priorities for disclosure decisions and outcomes specific to their age, race, relationship status, time since HIV diagnosis, and region. Participant feedback directed the development of the 4-module curriculum that includes general HIV information and communication skill-building activities. The main emphasis throughout the modular curriculum is practicing disclosure decisions, which gives users the ability to think through possible scenarios and create a personal strategy to approach disclosure

opportunities. This is intended to reduce the uncertainty and anxiety experienced prior to disclosing one's status. The RCT will determine whether it is more helpful to complete these modules at the user's own pace or in an environment where they can have technical or emotional support with the app. It will also address the timing of intervention, including whether the app is particularly useful for those recently diagnosed with HIV. This will help to guide the implementation of the app in the clinical setting in terms of timing and staffing.

AI has been successfully implemented in behavioral interventions for other chronic diseases [24,25], and the initial phases of TT confirmed its effectiveness for YMSM with HIV [17]. However, the emotional and psychological complexity of status disclosure requires additional development of the automated, interactive system used in TT. The crowdsourced disclosure scenarios from the comic book contest supplied additional content to the utterance database. This training data from the target population will enhance the simulated sex partner's dialogue capacity. The RCT will include a comparison of the acceptability and effectiveness of the AI (fully automated)– and wizard-of-oz (semiautomated)–supported dialogue platform.

### Limitations

Although participants reported that the virtual scenarios were reminiscent of their own disclosure experiences, texting input took away from the realistic nature of the conversation. Thus, the app was modified to enable the simulated sex partners to respond to spoken disclosure. Virtual reality interventions have been used for treating anxiety and other psychological conditions

[26], and previous studies have shown the emotional release of disclosing [27]. Subanalyses in the RCT will assess the potential therapeutic benefit of stating one's HIV status aloud. There were limitations to the comic book consent and usability study. The crowdsourcing method for the comic book allowed our team to reach a wider audience, but the participants did not indicate their HIV status. The usability study had a small sample from one geographic location, and all identified as Black or African American, which limits the generalizability of the findings. We plan to enroll a more diverse cohort for the RCT that includes a large proportion of people of color since HIV disproportionately affects populations with systemic health disparities. The usability study participants were recruited from a clinical setting, so many already had established care. These factors were taken into consideration in designing eligibility criteria for the RCT, which is the next phase for testing TT. The limitations to the use of the paradata will be noted during the RCT.

## Conclusion

Disclosure is not a one-time event, but a complex series of personal decisions and actions that will play out over the lifetime of youth with HIV. Despite advances in HIV care, HIV disclosure interventions that provide skills and information for YMSM living with HIV to make their own disclosure decisions are still needed regardless of time since diagnosis, viral suppression, or past disclosure experiences. TT was found to be an engaging and practical intervention for self-disclosure among YMSM living with HIV. Facilitating informed disclosure decisions has the potential to impact engagement in sexual risk behaviors and ART adherence. More information is needed about the ideal environment (eg, at home, in clinic) and technical and staff support required for implementation of an mHealth disclosure intervention. To that end, TT will be tested for effectiveness in a multisite RCT.

## Acknowledgments

This study was funded by grant 5R44MH104102-05 from National Institute of Mental Health to principal investigators MAL and LBHW.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Tough Talks technical pilot presurvey.

[[DOCX File, 77 KB - formative\\_v6i9e38354\\_app1.docx](#)]

## References

1. Chaudoir SR, Fisher JD, Simoni JM. Understanding HIV disclosure: a review and application of the Disclosure Processes Model. *Soc Sci Med* 2011 May;72(10):1618-1629 [FREE Full text] [doi: [10.1016/j.socscimed.2011.03.028](https://doi.org/10.1016/j.socscimed.2011.03.028)] [Medline: [21514708](https://pubmed.ncbi.nlm.nih.gov/21514708/)]
2. Thoth CA, Tucker C, Leahy M, Stewart SM. Self-disclosure of serostatus by youth who are HIV-positive: a review. *J Behav Med* 2014 Apr;37(2):276-288. [doi: [10.1007/s10865-012-9485-2](https://doi.org/10.1007/s10865-012-9485-2)] [Medline: [23277232](https://pubmed.ncbi.nlm.nih.gov/23277232/)]
3. Lee B, Oberdorfer P. Risk-taking behaviors among vertically HIV-infected adolescents in northern Thailand. *J Int Assoc Physicians AIDS Care (Chic)* 2009;8(4):221-228. [doi: [10.1177/1545109709341082](https://doi.org/10.1177/1545109709341082)] [Medline: [19596866](https://pubmed.ncbi.nlm.nih.gov/19596866/)]
4. Dempsey AG, MacDonell KE, Naar-King S, Lau C, Adolescent Medicine Trials Network for HIV/AIDS Interventions. Patterns of disclosure among youth who are HIV-positive: a multisite study. *J Adolesc Health* 2012 Mar;50(3):315-317 [FREE Full text] [doi: [10.1016/j.jadohealth.2011.06.003](https://doi.org/10.1016/j.jadohealth.2011.06.003)] [Medline: [22325140](https://pubmed.ncbi.nlm.nih.gov/22325140/)]
5. Cook CL, Staras SAS, Zhou Z, Chichetto N, Cook RL. Disclosure of HIV serostatus and condomless sex among men living with HIV/AIDS in Florida. *PLoS One* 2018;13(12):e0207838 [FREE Full text] [doi: [10.1371/journal.pone.0207838](https://doi.org/10.1371/journal.pone.0207838)] [Medline: [30557362](https://pubmed.ncbi.nlm.nih.gov/30557362/)]
6. Kalichman SC, Kalichman MO, Cherry C, Grebler T. HIV disclosure and transmission risks to sex partners among HIV-positive men. *AIDS Patient Care STDs* 2016 May;30(5):221-228 [FREE Full text] [doi: [10.1089/apc.2015.0333](https://doi.org/10.1089/apc.2015.0333)] [Medline: [27158850](https://pubmed.ncbi.nlm.nih.gov/27158850/)]
7. Fernet M, Wong K, Richard M, Otis J, Lévy JJ, Lapointe N, et al. Romantic relationships and sexual activities of the first generation of youth living with HIV since birth. *AIDS Care* 2011 Apr;23(4):393-400. [doi: [10.1080/09540121.2010.516332](https://doi.org/10.1080/09540121.2010.516332)] [Medline: [21271400](https://pubmed.ncbi.nlm.nih.gov/21271400/)]
8. Hightow-Weidman LB, Phillips G, Outlaw AY, Wohl AR, Fields S, Hildalgo J, et al. Patterns of HIV disclosure and condom use among HIV-infected young racial/ethnic minority men who have sex with men. *AIDS Behav* 2013 Jan;17(1):360-368. [doi: [10.1007/s10461-012-0331-x](https://doi.org/10.1007/s10461-012-0331-x)] [Medline: [23054043](https://pubmed.ncbi.nlm.nih.gov/23054043/)]
9. Mburu G, Hodgson I, Kalibala S, Haamujompa C, Cataldo F, Lowenthal ED, et al. Adolescent HIV disclosure in Zambia: barriers, facilitators and outcomes. *J Int AIDS Soc* 2014;17:18866 [FREE Full text] [doi: [10.7448/IAS.17.1.18866](https://doi.org/10.7448/IAS.17.1.18866)] [Medline: [24629845](https://pubmed.ncbi.nlm.nih.gov/24629845/)]

10. Simoni JM, Montgomery A, Martin E, New M, Demas PA, Rana S. Adherence to antiretroviral therapy for pediatric HIV infection: a qualitative systematic review with recommendations for research and clinical management. *Pediatrics* 2007 Jun;119(6):e1371-e1383 [FREE Full text] [doi: [10.1542/peds.2006-1232](https://doi.org/10.1542/peds.2006-1232)] [Medline: [17533177](https://pubmed.ncbi.nlm.nih.gov/17533177/)]
11. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Antiretroviral therapy for the prevention of HIV-1 transmission. *N Engl J Med* 2016 Sep 01;375(9):830-839 [FREE Full text] [doi: [10.1056/NEJMoa1600693](https://doi.org/10.1056/NEJMoa1600693)] [Medline: [27424812](https://pubmed.ncbi.nlm.nih.gov/27424812/)]
12. Gabbidon K, Chenneville T, Peless T, Sheared-Evans S. Self-disclosure of HIV status among youth living with HIV: a global systematic review. *AIDS Behav* 2020 Jan;24(1):114-141. [doi: [10.1007/s10461-019-02478-9](https://doi.org/10.1007/s10461-019-02478-9)] [Medline: [30924065](https://pubmed.ncbi.nlm.nih.gov/30924065/)]
13. Estimated HIV incidence and prevalence in the United States, 2010-2016. HIV Surveillance Supplemental Report. Atlanta: Centers for Disease Control and Prevention; 2019. URL: <https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-24-1.pdf> [accessed 2022-07-27]
14. Garofalo R, Hottot A, Kuhns L, Gratz B, Mustanski B. Incidence of HIV infection and sexually transmitted infections and related risk factors among very young men who have sex with men. *J Acquir Immune Defic Syndr* 2016 May 01;72(1):79-86 [FREE Full text] [doi: [10.1097/QAI.0000000000000933](https://doi.org/10.1097/QAI.0000000000000933)] [Medline: [26745827](https://pubmed.ncbi.nlm.nih.gov/26745827/)]
15. Zandoni BC, Mayer KH. The adolescent and young adult HIV cascade of care in the United States: exaggerated health disparities. *AIDS Patient Care STDS* 2014 Mar;28(3):128-135 [FREE Full text] [doi: [10.1089/apc.2013.0345](https://doi.org/10.1089/apc.2013.0345)] [Medline: [24601734](https://pubmed.ncbi.nlm.nih.gov/24601734/)]
16. Cao W, Wong HM, Chang C, Agudile EP, Ekström AM. Behavioral interventions promoting HIV serostatus disclosure to sex partners among HIV-positive men who have sex with men: a systematic review. *Int J Public Health* 2019 Sep;64(7):985-998. [doi: [10.1007/s00038-019-01275-4](https://doi.org/10.1007/s00038-019-01275-4)] [Medline: [31250027](https://pubmed.ncbi.nlm.nih.gov/31250027/)]
17. Muessig KE, Knudtson KA, Soni K, Larsen MA, Traum D, Dong W, et al. "I didn't tell you sooner because I didn't know how to handle it myself." Developing a virtual reality program to support HIV-status disclosure decisions. *Digit Cult Educ* 2018;10:22-48 [FREE Full text] [Medline: [30123342](https://pubmed.ncbi.nlm.nih.gov/30123342/)]
18. Hightow-Weidman LB, Bauermeister JA. Engagement in mHealth behavioral interventions for HIV prevention and care: making sense of the metrics. *Mhealth* 2020;6:7 [FREE Full text] [doi: [10.21037/mhealth.2019.10.01](https://doi.org/10.21037/mhealth.2019.10.01)] [Medline: [32190618](https://pubmed.ncbi.nlm.nih.gov/32190618/)]
19. Bauermeister JA, Golinkoff JM, Muessig KE, Horvath KJ, Hightow-Weidman LB. Addressing engagement in technology-based behavioural HIV interventions through paradata metrics. *Curr Opin HIV AIDS* 2017 Sep;12(5):442-446 [FREE Full text] [doi: [10.1097/COH.0000000000000396](https://doi.org/10.1097/COH.0000000000000396)] [Medline: [28617711](https://pubmed.ncbi.nlm.nih.gov/28617711/)]
20. Bandura A. Health promotion from the perspective of social cognitive theory. *Psychology & Health* 1998 Jul;13(4):623-649. [doi: [10.1080/08870449808407422](https://doi.org/10.1080/08870449808407422)]
21. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. *Epidemiol Psychiatr Sci* 2016 Apr;25(2):113-122 [FREE Full text] [doi: [10.1017/S2045796015001067](https://doi.org/10.1017/S2045796015001067)] [Medline: [26744309](https://pubmed.ncbi.nlm.nih.gov/26744309/)]
22. Arnold EA, Rebchook GM, Kegeles SM. 'Triply cursed': racism, homophobia and HIV-related stigma are barriers to regular HIV testing, treatment adherence and disclosure among young Black gay men. *Cult Health Sex* 2014 Jun;16(6):710-722 [FREE Full text] [doi: [10.1080/13691058.2014.905706](https://doi.org/10.1080/13691058.2014.905706)] [Medline: [24784224](https://pubmed.ncbi.nlm.nih.gov/24784224/)]
23. Maiorana A, Kegeles SM, Brown S, Williams R, Arnold EA. Substance use, intimate partner violence, history of incarceration and vulnerability to HIV among young Black men who have sex with men in a Southern US city. *Cult Health Sex* 2021 Jan;23(1):37-51 [FREE Full text] [doi: [10.1080/13691058.2019.1688395](https://doi.org/10.1080/13691058.2019.1688395)] [Medline: [31944158](https://pubmed.ncbi.nlm.nih.gov/31944158/)]
24. Ahn SJ. Incorporating immersive virtual environments in health promotion campaigns: a construal level theory approach. *Health Commun* 2015;30(6):545-556. [doi: [10.1080/10410236.2013.869650](https://doi.org/10.1080/10410236.2013.869650)] [Medline: [24991725](https://pubmed.ncbi.nlm.nih.gov/24991725/)]
25. Morie JF, Chance E. Extending the reach of health care for obesity and diabetes using virtual worlds. *J Diabetes Sci Technol* 2011 Mar 01;5(2):272-276 [FREE Full text] [doi: [10.1177/193229681100500211](https://doi.org/10.1177/193229681100500211)] [Medline: [21527093](https://pubmed.ncbi.nlm.nih.gov/21527093/)]
26. Maples-Keller JL, Bunnell BE, Kim S, Rothbaum BO. The Use of Virtual Reality Technology in the Treatment of Anxiety and Other Psychiatric Disorders. *Harv Rev Psychiatry* 2017;25(3):103-113. [doi: [10.1097/hrp.0000000000000138](https://doi.org/10.1097/hrp.0000000000000138)]
27. Serovich JM, Mosack KE. Reasons for HIV disclosure or nondisclosure to casual sexual partners. *AIDS Educ Prev* 2003 Feb;15(1):70-80 [FREE Full text] [doi: [10.1521/aeap.15.1.70.23846](https://doi.org/10.1521/aeap.15.1.70.23846)] [Medline: [12627744](https://pubmed.ncbi.nlm.nih.gov/12627744/)]

## Abbreviations

- AI:** artificial intelligence
- ART:** antiretroviral therapy
- BATLab:** Behavior and Technology Lab
- CAI:** condomless anal intercourse
- mHealth:** mobile health
- MSM:** men who have sex with men
- RCT:** randomized control trial
- SCT:** social cognitive theory
- TT:** Tough Talks

**UNC-CH:** University of North Carolina at Chapel Hill

**VBI:** Virtually Better, Inc

**YMSM:** young MSM

*Edited by A Mavragani; submitted 29.03.22; peer-reviewed by G Sabben, C Lelutiu-Weinberger; comments to author 09.06.22; revised version received 25.07.22; accepted 25.07.22; published 08.09.22.*

*Please cite as:*

*Hightow-Weidman LB, Muessig K, Soberano Z, Rosso MT, Currie A, Adams Larsen M, Knudtson K, Vecchio A  
Tough Talks Virtual Simulation HIV Disclosure Intervention for Young Men Who Have Sex With Men: Development and Usability  
Testing*

*JMIR Form Res 2022;6(9):e38354*

*URL: <https://formative.jmir.org/2022/9/e38354>*

*doi: [10.2196/38354](https://doi.org/10.2196/38354)*

*PMID: [36074551](https://pubmed.ncbi.nlm.nih.gov/36074551/)*

©Lisa B Hightow-Weidman, Kathryn Muessig, Zach Soberano, Matthew T Rosso, Andrew Currie, Margo Adams Larsen, Kelly Knudtson, Alyssa Vecchio. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 08.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Optimizing Health Coaching for Patients With Type 2 Diabetes Using Machine Learning: Model Development and Validation Study

Shuang Di<sup>1,2</sup>, BSc, MSc; Jeremy Petch<sup>1,3,4,5</sup>, MA, HBA, PhD; Hertz C Gerstein<sup>3,5</sup>, MSc, MD, FRCPC; Ruoqing Zhu<sup>6</sup>, PhD; Diana Sherifali<sup>3,7</sup>, PhD, RN, CDE

<sup>1</sup>Centre for Data Science and Digital Health, Hamilton Health Sciences, Hamilton, ON, Canada

<sup>2</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

<sup>3</sup>Population Health Research Institute, Hamilton Health Sciences, Hamilton, ON, Canada

<sup>4</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

<sup>5</sup>Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

<sup>6</sup>Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, United States

<sup>7</sup>School of Nursing, McMaster University, Hamilton, ON, Canada

**Corresponding Author:**

Diana Sherifali, PhD, RN, CDE

School of Nursing

McMaster University

1280 Main Street West

Hamilton, ON, L8S 4K1

Canada

Phone: 1 9055259140 ext 21435

Email: [dsherif@mcmaster.ca](mailto:dsherif@mcmaster.ca)

## Abstract

**Background:** Health coaching is an emerging intervention that has been shown to improve clinical and patient-relevant outcomes for type 2 diabetes. Advances in artificial intelligence may provide an avenue for developing a more personalized, adaptive, and cost-effective approach to diabetes health coaching.

**Objective:** We aim to apply Q-learning, a widely used reinforcement learning algorithm, to a diabetes health-coaching data set to develop a model for recommending an optimal coaching intervention at each decision point that is tailored to a patient's accumulated history.

**Methods:** In this pilot study, we fit a two-stage reinforcement learning model on 177 patients from the intervention arm of a community-based randomized controlled trial conducted in Canada. The policy produced by the reinforcement learning model can recommend a coaching intervention at each decision point that is tailored to a patient's accumulated history and is expected to maximize the composite clinical outcome of hemoglobin A<sub>1c</sub> reduction and quality of life improvement (normalized to [0, 1], with a higher score being better). Our data, models, and source code are publicly available.

**Results:** Among the 177 patients, the coaching intervention recommended by our policy mirrored the observed diabetes health coach's interventions in 17.5% (n=31) of the patients in stage 1 and 14.1% (n=25) of the patients in stage 2. Where there was agreement in both stages, the average cumulative composite outcome (0.839, 95% CI 0.460-1.220) was better than those for whom the optimal policy agreed with the diabetes health coach in only one stage (0.791, 95% CI 0.747-0.836) or differed in both stages (0.755, 95% CI 0.728-0.781). Additionally, the average cumulative composite outcome predicted for the policy's recommendations was significantly better than that of the observed diabetes health coach's recommendations ( $t_{n-1}=10.040$ ;  $P<.001$ ).

**Conclusions:** Applying reinforcement learning to diabetes health coaching could allow for both the automation of health coaching and an improvement in health outcomes produced by this type of intervention.

(JMIR Form Res 2022;6(9):e37838) doi:[10.2196/37838](https://doi.org/10.2196/37838)

**KEYWORDS**

diabetes health coaching; artificial intelligence; reinforcement learning; health coaching; patient outcome; diabetes; community health; digital intervention; health outcome

## *Introduction*

Chronic diseases are a major health care challenge globally and domestically, being the leading cause of death and disability worldwide as of 2021 [1,2] and accounting for 89% of all deaths in Canada [3]. As of 2011, type 2 diabetes (T2D) affects more than 2.5 million people in Canada specifically and costs the health care system over CAD \$6.7 billion (US \$5.1 billion) to treat annually [4].

Health coaching is quickly emerging as a new approach to partner with patients to optimize their self-management through lifestyle changes [5]. Diabetes health coaching has both educational and behavioral components, which include goal-setting, self-care knowledge, and frequent follow-up appointments [6]. Coaching has been shown to improve health outcomes [7-9] and treatment adherence [10,11]. However, the widespread adoption of diabetes health coaching may be limited by constraints on health human resources. Artificial intelligence incorporated into a digital health platform could automate some routine health-coaching tasks to improve the scalability of coaching interventions. Moreover, artificial intelligence may be able to leverage data from a patient's history that is not routinely used in clinical practice to optimize coaching recommendations.

Recent work in artificial intelligence and medicine [12] suggests that individual patient data can be leveraged to assist the decision-making process of diabetes health coaching and suggests incremental adjustments of interventions tailored to the patient's changing needs and health status. Reinforcement learning is commonly used for estimating an optimal set of actions (called a "policy") for this type of sequential decision-making problem [13,14]. Reinforcement learning works by iteratively choosing actions and, then in turn, is rewarded based on the outcomes of those actions. This is done for every

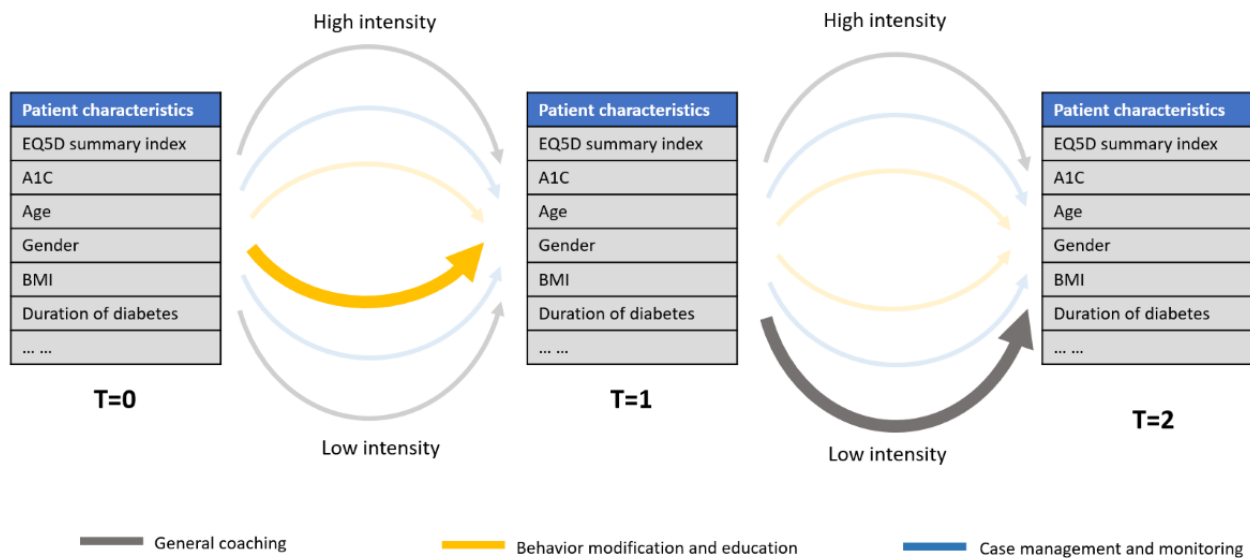
set of patient characteristics at every time step in a data set. The algorithm "learns" the best action to take at every time step by maximizing the value of the rewards over all time steps for each patient (Figure 1).

Several studies have applied reinforcement learning for diabetes management, with most focused on controlling blood glucose levels [15,16]. Vrabie et al [17] used reinforcement learning to obtain optimal adaptive control algorithms for dynamical systems using mathematical models. Ngo et al [18,19] applied reinforcement learning for the optimal control of blood glucose in patients with type 1 diabetes and proposed a reinforcement learning algorithm for automatically calculating the basal and bolus insulin doses for patients with diabetes using a simulation on a blood glucose model.

Relatively few studies have used reinforcement learning for diabetes health coaching [20-22]. Yom-Tov et al [20] developed a reinforcement learning-powered system that used personalized messages to improve T2D patients' compliance with their physical activity regimens. Lauffenburger et al [22] have developed a reinforcement learning-powered system to personalize SMS text messages to promote medication adherence. This existing reinforcement learning research for diabetes health coaching focuses on improving coaching within a single domain (only physical activities or only medication adherence). Our study is the first to use reinforcement learning to recommend comprehensive coaching strategies that can include all domains of coaching (physical activity, medication adherence, diet modification, etc) that are tailored based on patients' changing clinical status and ongoing performance.

In this study, we applied Q-learning (Multimedia Appendix 1), a widely used reinforcement learning algorithm, to a diabetes health-coaching data set to develop a model for recommending an optimal coaching intervention at each decision point that is tailored to a patient's accumulated history.

**Figure 1.** Diabetes health-coaching optimization as a sequential decision-making problem. A<sub>1c</sub>: Hemoglobin A<sub>1c</sub>; ED5D: EuroQol five-dimension scale questionnaire; T: time step.



## Methods

### Data Overview

The data set used in this study was collected in a community-based randomized controlled trial conducted in Ontario, Canada [23]. Patients in the trial were 18 years or older, diagnosed with T2D (any duration), and had a hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>) level >7.5% within 6 months prior to randomization. All patients were able to read and write in English, and had access to a telephone. Those excluded were pregnant women, had debilitating coexisting conditions (ie, mental illness or impaired cognition), or had underlying medical conditions that may have provided misleading HbA<sub>1c</sub> levels. A total of 365 patients were randomized using a 1:1 ratio into the intervention (diabetes health coaching) or control (usual care) groups. All patients in the intervention arm of the trial were included in the current analysis.

Patients in the intervention arm received both standard care and an additional diabetes health-coaching intervention. Standard care included receiving access to usual diabetes education (individual or group) provided by nurses or dietitians, typically every 3 to 6 months, along with community resources. In addition, the intervention group received diabetes health coaching delivered by a registered nurse or certified diabetes educator that emphasized small positive habits customized to one's environment, ability, and motivation. The topic or agenda of each telephone call was determined by the participant or was agreed upon in the previous coaching session. All patients in the intervention arm had access to diabetes health coaching for 1 year.

For each patient, the data set contained demographic data, including age, gender, ethnicity, diabetes duration, and comorbidities; clinical characteristics, including BMI, weight, and most recent HbA<sub>1c</sub>; health care resource use information, including hospital admissions, emergency room visits, specialist visits, and other health care visits (eg, nurse visits); and quality

of life (QoL) measures. Demographic data and health care resource use were collected using self-reported questionnaires, and clinical characteristics were assessed at study visits or through medical records. QoL was measured using three scales, including the Audit of Diabetes-Dependent Quality of Life (ADDQoL) scale [24], the Diabetes Self-Care Activities (DSCA) scale [25], and the EQ-5D scale [26]. All the measures were collected at baseline and at the 6-month and 12-month follow-ups. A coaching intervention use form was used to document the diabetes health coaching received by each patient over the course of the trial. A patient could visit the diabetes health coach multiple times during the trial and could receive one or multiple coaching recommendations at each visit: dietary modification, exercise modification, behavioral modification, medication adherence, medication adjustment, glucose monitoring, psychological support or counseling, case management/monitoring, and system navigation.

We have made the data set used in this study publicly available [27].

### Ethical Approval

The trial from which our data set was derived was approved by the Hamilton Integrated Research Ethics Board (approval/file number: 14-416). Written informed consent was obtained from all participants (inability to provide informed consent was an exclusion criteria for the trial) and included permission for a secondary analysis without additional consent. The trial was registered at ClinicalTrials.gov (NCT02128815) [28]. All data for the trial was deidentified. Participants were provided a small honorarium of CAD \$20 (US \$15.25) per visit for over three visits (thus, a total of CAD \$60 [US \$45.75]) for participation in the trial.

### Problem Formulation for the Reinforcement Learning Model

Reinforcement learning is an approach to machine learning inspired by how animals and humans can learn new tasks through receiving rewards for desirable behavior. For example,

dogs are often taught to perform tricks by giving them treats after performing well. In reinforcement learning, an algorithm (referred to as an “agent”) learns an optimal policy through trial and error within a simulated environment. During the learning process, the agent will make decisions based on inputs from the environment and then will receive rewards if those decisions resulted in a desirable outcome. Over many iterations, the agent eventually learns an optimal strategy (referred to as a “policy”) that allows it to consistently maximize rewards.

In this study, our goal was to use reinforcement learning to learn an optimal policy for recommending diabetes coaching interventions at each clinical decision point, using a patient’s accumulated history as inputs. We rewarded the agent based on a composite outcome of HbA<sub>1c</sub> reduction and QoL improvement (measured using the EQ-5D summary index, which was chosen based on expert clinical input). We set both weights to 0.5 to reflect equal importance and additionally scaled both HbA<sub>1c</sub> reduction and QoL improvements to the range of [0, 1] before calculating the weighted average. For example, 1 patient had an HbA<sub>1c</sub> of 7.0 and an EQ-5D summary index of 0.457 at baseline, and then at the 6-month follow-up, their HbA<sub>1c</sub> decreased to 6.8 and their EQ-5D summary index increased to 0.533. We calculated their reduction in HbA<sub>1c</sub> as 0.835 (2.86% reduction before min-max scaling) and their increase in QoL as 0.504 (16.5% improvement before min-max scaling). The weighted average was calculated as  $0.5 * 0.835 + 0.5 * 0.504$ , which is 0.670. The reinforcement learning agent was rewarded based on the cumulative composite outcome, which we calculated by adding together the composite outcome as recorded at both the 6-month and 12-month follow-ups.

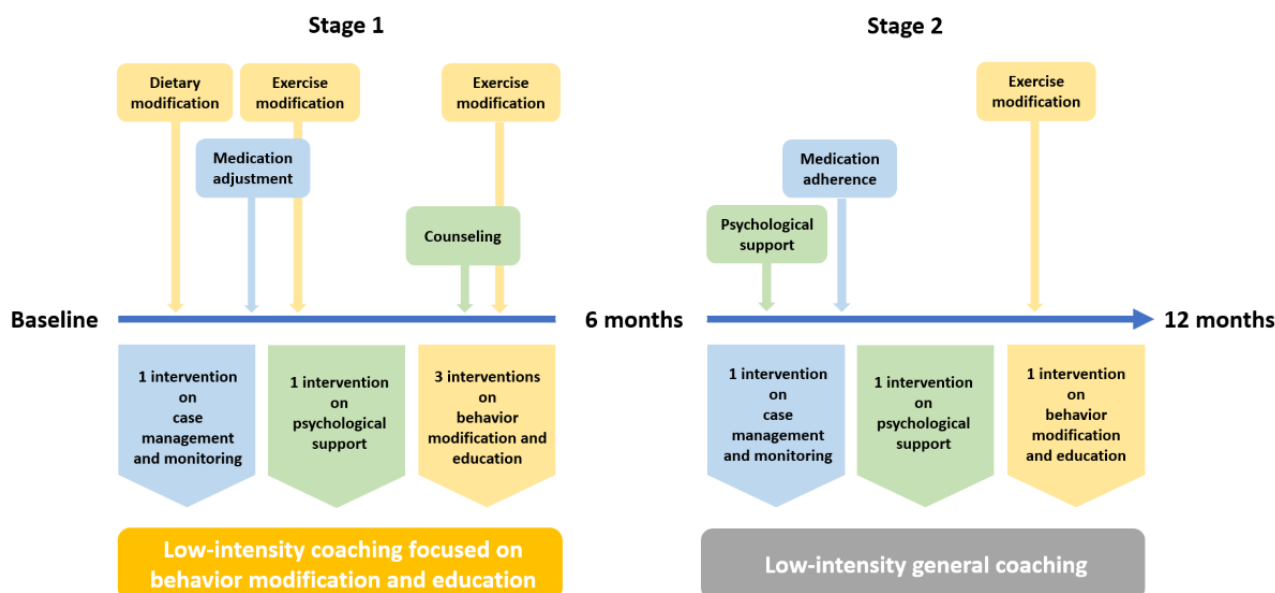
To prepare the simulated environment necessary for reinforcement learning, we first identified the patient characteristics, decision points, and intervention options from the data set. Patient characteristics included demographic data, clinical characteristics, health care resource use information,

and self-reported QoL measured by the ADDQoL and DSCA scales. Since we had access to patient characteristics and the outcome of interest measured at baseline, the 6 month follow-up, and the 12-month follow-up, we formalized the sequence of data as a 2-stage estimation problem, with the 2 decision points being the initial visit and the 6-month follow-up.

In reinforcement learning, the available options for interventions are called the “action space.” Action spaces that are too complex can cause challenges with the learning process, so it is standard practice in reinforcement learning to “shape” the action space by constraining the available options in some way—often by combining very similar actions [29]. For our study, we grouped 9 distinct diabetes coaching recommendations into 3 categories based on expert clinical input. Specifically, dietary modification, exercise modification, and behavioral modification were grouped into the category of behavior modification and education; medication adherence, medication adjustment, glucose monitoring, case management/monitoring, and system navigation were grouped into the category of case management and monitoring; and psychological support and counseling were combined into the category of psychological support. To be classified as one of these 3 categories, a patient needed to have at least twice as many recommendations in 1 category compared to the others—otherwise they were classified as a fourth category: general coaching.

In addition to shaping the action space based on the focus of the interventions, we also categorized interventions based on intensity. We categorized intensity by calculating the total number of coaching recommendations received by each patient in a stage to obtain the median of the total number of coaching recommendations among all patients. High-intensity coaching was categorized as being greater than the median number of coaching recommendations during a time interval, and low-intensity coaching was categorized as fewer than the median. The dimensions of focus and intensity resulted in an action space with 8 possible actions (Figure 2).

Figure 2. An example of a single patient in the data set.





## Optimal Policy Estimation and Validation

Following problem formulation, we fit a reinforcement learning model to “learn” which interventions tend to produce the best outcome for each set of patient characteristics. The Q-learning algorithm formulates this as a prediction problem, with patient characteristics and coaching actions as model inputs used to predict the cumulative composite outcome, which we use as the reward function. This prediction model is then used to select the optimal action for a given set of patient characteristics by estimating the rewards for all possible actions from the action space and selecting the one estimated to produce the greatest reward. While any regression modeling technique can be used for this type of prediction problem, we selected histogram-based gradient boosting classification trees [30], as they are better suited to modeling large numbers of patient characteristics with complex interactions than techniques like linear regression.

Given the relatively small sample size available for this pilot, we were able to use a leave-one-out cross-validation (LOOCV) approach [31] for model development and validation. This approach trains a model on all available patients but 1, then uses the remaining patient for model validation. This process is then repeated for every possible split of the data, resulting in 177 iterations of train/test for our data set. This hypothesis was tested using a paired  $t$  test.

The potential clinical effectiveness of our model was evaluated using two approaches. The first approach compared the model predicted cumulative composite outcome with the actual observed outcome (the observed result of the interventions provided by the diabetes health coaches). We hypothesized that our model’s predicted outcome would be higher than the observed outcome.

The second approach assessed the relationship between the cumulative composite outcome and the proportion of agreement between our model and the diabetes health coach. We hypothesized that higher levels of agreement between our model and coach recommendations would be associated with better observed outcomes and that lower levels of agreement would be associated with worse outcomes.

The level of agreement between the reinforcement learning model and the observed diabetes health coach’s interventions was not used to evaluate the performance of our model. This is because reinforcement learning assumes that there is room for improvement over observed behavior. Thus the goal is to learn a different better policy than what was observed, rather than simply mirroring what was done by the coaches.

We have made the source code and trained models developed for this study publicly available [27].

## Results

A total of 177 patients in the intervention arm of the community-based randomized controlled trial were included in the analysis. Distributions of the patient characteristics used as model inputs at baseline and the 6-month and 12-month follow-ups are summarized in Table 1.  $P$  values are reported to illustrate the degree of change for each characteristic between time points.

The median of the total number of coaching recommendations received in stage 1 and stage 2 was 8. Following the prespecified criteria for defining intervention options, we obtained the following intervention options (Table 2).

Among the 177 patients, LOOCV results showed that the average cumulative composite outcome expected by the reinforcement learning model (0.811) was significantly higher than the observed outcome (0.767;  $t_{n-1}=10.040$ ;  $P<.001$ ).

LOOCV results also showed that our model mirrored the observed diabetes health coach’s interventions in 17.5% ( $n=31$ ) of the patients in stage 1 and in 14.1% ( $n=25$ ) of the patients in stage 2. Among the patients for whom our model agreed with the diabetes health coach in both stages, the average cumulative composite outcome (0.839, 95% CI 0.460-1.220) was better than those for whom our model agreed with the diabetes health coach in only one stage (0.791, 95% CI 0.747-0.836) or differed in both stages (0.755, 95% CI 0.728-0.781).

**Table 1.** Patient characteristics used as model inputs with SDs and percentages at baseline, the 6-month follow-up, and the 12-month follow-up (N=177).

Patient characteristics	Baseline	6-month follow-up	12-month follow-up	P value
Age (years), mean (SD)	57.4 (11.3)	57.4 (11.3)	57.4 (11.3)	N/A <sup>a</sup>
<b>Gender, n (%)</b>				N/A
Female	94 (53.1)	94 (53.1)	94 (53.1)	
Male	83 (46.9)	83 (46.9)	83 (46.9)	
<b>Ethnicity, n (%)</b>				N/A
Caucasian	141 (79.7)	141 (79.7)	141 (79.7)	
Latin American	10 (5.6)	10 (5.6)	10 (5.6)	
South Asian	10 (5.6)	10 (5.6)	10 (5.6)	
Aboriginal	3 (1.7)	3 (1.7)	3 (1.7)	
Filipino	3 (1.7)	3 (1.7)	3 (1.7)	
Black	2 (1.1)	2 (1.1)	2 (1.1)	
Southeast Asian	2 (1.1)	2 (1.1)	2 (1.1)	
Arab	1 (0.6)	1 (0.6)	1 (0.6)	
Chinese	1 (0.6)	1 (0.6)	1 (0.6)	
West Asian	1 (0.6)	1 (0.6)	1 (0.6)	
Unknown	3 (1.7)	3 (1.7)	3 (1.7)	
BMI, mean (SD)	34.5 (6.9)	34.1 (7.2)	33.6 (6.9)	.51
Duration of diabetes (years), mean (SD)	9.4 (9.1)	9.9 (9.1)	10.4 (9.1)	.60
Hemoglobin A <sub>1c</sub> , mean (SD)	9.1 (1.7)	7.6 (1.2)	7.3 (1.1)	<.001
Family physician visits, mean (SD)	3.0 (2.9)	2.4 (2.3)	2.2 (1.9)	.003
Family physician visits related to diabetes, mean (SD)	1.7 (1.6)	1.6 (1.6)	1.4 (0.9)	.11
Visits with health professional, n (%)	45 (25.4)	55 (31.1)	80 (45.2)	<.001
Emergency room and hospital admissions, n (%)	156 (88.1)	160 (90.4)	116 (65.5)	<.001
Chronic disease management program, n (%)	164 (92.7)	173 (97.7)	126 (71.2)	<.001
<b>Behavioral stage, n (%)</b>				<.001
Action	103 (58.2)	128 (72.3)	104 (58.8)	
Contemplation	15 (8.5)	6 (3.4)	3 (1.7)	
I am not sure	2 (1.1)	1 (0.6)	0 (0.0)	
Maintenance	0 (0.0)	24 (13.6)	52 (29.4)	
Precontemplation	7 (4.0)	0 (0.0)	0 (0.0)	
Preparation	50 (28.2)	18 (10.2)	18 (10.2)	
<b>Diabetes treatment, n (%)</b>				
Diet	85 (48.0)	70 (39.5)	43 (24.3)	<.001
Oral therapy	163 (92.1)	166 (93.8)	165 (93.2)	.82
Insulin	70 (39.5)	77 (43.5)	76 (42.9)	.72
Other	3 (1.7)	2 (1.1)	0 (0.0)	.24
EQ-5D summary index, mean (SD)	0.8 (0.2)	0.8 (0.1)	0.8 (0.1)	.06
ADDQoL <sup>b</sup> summary score, mean (SD)	-1.5 (1.3)	-1.4 (1.1)	-1.3 (0.7)	.10
<b>Diabetes Self-Care Activities scale</b>				
General diet, mean (SD)	4.5 (2.8)	5.6 (2.3)	6.1 (1.9)	<.001
Specific diet, mean (SD)	5.1 (1.6)	5.4 (1.3)	5.6 (1.1)	.003

Patient characteristics	Baseline	6-month follow-up	12-month follow-up	P value
Exercise, mean (SD)	4.1 (2.6)	5.3 (2.4)	5.6 (2.3)	<.001
Blood glucose testing, mean (SD)	5.3 (2.4)	5.8 (2.0)	5.5 (2.2)	.13
Foot care, mean (SD)	2.9 (1.9)	3.5 (1.3)	2.9 (1.7)	<.001
Current smoker, n (%)	29 (16.4)	24 (13.6)	23 (13.0)	.62
Cigarettes smoked per day, mean (SD)	2.8 (7.2)	2.3 (7.2)	2.0 (6.8)	.62
Additional diet <sup>c</sup> , mean (SD)	3.7 (3.2)	4.9 (3.1)	5.3 (2.9)	<.001
Additional medication, mean (SD)	6.7 (1.4)	6.9 (0.9)	6.9 (0.9)	.18
Additional foot care, mean (SD)	5.9 (1.4)	6.4 (1.1)	6.5 (0.8)	<.001
Stroke, n (%)	5 (2.8)	0 (0.0)	1 (0.6)	.03
Transient ischemic attack, n (%)	17 (9.6)	1 (0.6)	0 (0.0)	<.001
Evidence of coronary artery disease, n (%)	17 (9.6)	0 (0.0)	0 (0.0)	<.001
Myocardial infarction, n (%)	4 (2.3)	0 (0.0)	0 (0.0)	.02
Heart failure, n (%)	0 (0.0)	0 (0.0)	1 (0.6)	.37
Kidney disease, n (%)	10 (5.6)	1 (0.6)	0 (0.0)	<.001
Chronic obstructive pulmonary disease, n (%)	19 (10.7)	3 (1.7)	3 (1.7)	<.001
Hyperlipidemia, n (%)	94 (53.1)	6 (3.4)	0 (0.0)	<.001
Hypertension, n (%)	108 (61.0)	9 (5.1)	0 (0.0)	<.001
Peripheral arterial disease, n (%)	3 (1.7)	2 (1.1)	0 (0.0)	.24
Prescribed medications, n (%)	5 (2.8)	45 (25.4)	29 (16.4)	<.001

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

<sup>b</sup>ADDQoL: Audit of Diabetes-Dependent Quality of Life.

<sup>c</sup>Additional items for the expanded version of the summary of Diabetes Self-Care Activities.

**Table 2.** Intervention options.

Intervention	Coaching recommendations (stage 1), n	Coaching recommendations (stage 2), n
High-intensity general coaching	45	18
High-intensity coaching on case management and monitoring	34	18
High-intensity coaching on behavior modification and education	29	10
Low-intensity general coaching	23	64
Low-intensity coaching on case management and monitoring	30	42
Low-intensity coaching on behavior modification and education	16	25

## Discussion

The study took a novel approach of developing artificial intelligence using diabetes health-coaching data to better fit the needs of diabetes management and to achieve better health outcomes. Using historical observational data from a community-based randomized controlled trial, we developed a reinforcement learning model that can automate the task of personalized adaptive diabetes health coaching and demonstrates the potential to outperform human diabetes health coaches in maximizing a composite outcome of HbA<sub>1c</sub> reduction and QoL improvement. Our approach is also able to leverage data that is

often overlooked, such as self-reported behavioral data, which allows us to generate personalized adaptive interventions for each patient using comprehensive health data.

The model-based decision-making process is fully automated, which requires less involvement from health care professional resources. In practice, our model could be integrated into existing diabetes health-coaching programs to dynamically suggest personalized adaptive coaching interventions, either as a decision-making support tool for the diabetes health coaches or combined with a patient-facing mobile app to directly support patients with diabetes, which has the potential to reduce the cost and expand the reach of diabetes health coaching [32,33].

This study has several limitations. The internal working of the reinforcement learning model is difficult to interpret, and as a result, the model appears as a black box to health care professionals and patients, which may present a barrier to adoption in some clinical settings [34]. Due to the relatively small sample size, the data source for this study lacks heterogeneity, which may result in insufficient generalizability of the estimated optimal policy, despite its satisfactory performance on the study population. We plan to address this limitation in future work, which will seek to include a larger and more diverse group of patients. The aggregation of detailed diabetes health-coaching data into discrete intervention options may have led to a loss of fidelity, which in turn may translate into less optimal intervention recommendations. Future work in this area may look to more advanced statistical methods to fully use the fine-grained original coaching information to

produce a better performance. Finally, diabetes health coach's interventions can potentially have different consequences on patients due to the human factors (eg, patients' adherence to coaching) that cannot be fully simulated, which may lead to lower performance in real-world clinical practice. Future work should investigate quantifying these human factors and including them in the reinforcement learning model.

This pilot study presents a novel application of artificial intelligence in diabetes management and demonstrated that applying reinforcement learning to diabetes health-coaching data has the potential to automate coaching and yield substantial improvement in health outcomes. Future research will include applying the reinforcement learning approach to larger diabetes health-coaching data sets and exploring the feasibility and acceptability of diabetes health coaching supported by artificial intelligence.

---

## Acknowledgments

This study was funded by Hamilton Health Sciences.

---

## Conflicts of Interest

HCG holds the McMaster-Sanofi Population Health Institute Chair in Diabetes Research and Care. He reports research grants from Eli Lilly, AstraZeneca, Merck, Novo Nordisk, and Sanofi; honoraria for speaking from Eli Lilly, Novo Nordisk, Sanofi, DKSH, Roche, and Zuellig; and consulting fees from Abbott, Covance, Eli Lilly, Novo Nordisk, Sanofi, Pfizer, Kowa, and Hanmi.

---

Multimedia Appendix 1

Q-learning.

[[DOCX File, 16 KB - formative\\_v6i9e37838\\_app1.docx](#)]

---

## References

1. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018 Feb;14(2):88-98. [doi: [10.1038/nrendo.2017.151](#)] [Medline: [29219149](#)]
2. Mazzucca S, Arredondo EM, Hoelscher DM, Haire-Joshu D, Tabak RG, Kumanyika SK, et al. Expanding implementation research to prevent chronic diseases in community settings. *Annu Rev Public Health* 2021 Apr 01;42:135-158 [[FREE Full text](#)] [doi: [10.1146/annurev-publhealth-090419-102547](#)] [Medline: [33467924](#)]
3. Wong-Rieger D. Health coaching for chronic conditions engaging and supporting patients to self manage. *7Acoach*. URL: [http://www.7acoach.com/PDF/Article\\_D\\_wong.pdf](http://www.7acoach.com/PDF/Article_D_wong.pdf) [accessed 2022-09-08]
4. Doucet G, Beatty M. The cost of diabetes in Canada: the economic tsunami. *Can J Diabetes* 2010 Jan;34(1):27-29. [doi: [10.1016/s1499-2671\(10\)41005-9](#)]
5. Wolever R, Simmons L, Sforzo G, Dill D, Kaye M, Bechard E, et al. A systematic review of the literature on health and wellness coaching: defining a key behavioral intervention in healthcare. *Glob Adv Health Med* 2013 Jul;2(4):38-57 [[FREE Full text](#)] [doi: [10.7453/gahmj.2013.042](#)] [Medline: [24416684](#)]
6. Sherifali D. Diabetes coaching for individuals with type 2 diabetes: a state-of-the-science review and rationale for a coaching model. *J Diabetes* 2017 Jun;9(6):547-554. [doi: [10.1111/1753-0407.12528](#)] [Medline: [28084681](#)]
7. Pirbaglou M, Katz J, Motamed M, Pludwinski S, Walker K, Ritvo P. Personal health coaching as a type 2 diabetes mellitus self-management strategy: a systematic review and meta-analysis of randomized controlled trials. *Am J Health Promot* 2018 Sep;32(7):1613-1626. [doi: [10.1177/0890117118758234](#)] [Medline: [29658286](#)]
8. Sherifali D, Viscardi V, Bai J, Ali RMU. Evaluating the effect of a diabetes health coach in individuals with type 2 diabetes. *Can J Diabetes* 2016 Feb;40(1):84-94. [doi: [10.1016/j.jcjd.2015.10.006](#)] [Medline: [26827684](#)]
9. Sherifali D, Brozic A, Agema P, Punthakee Z, McInnes N, O'Reilly D, et al. Effect of diabetes health coaching on glycemic control and quality of life in adults living with type 2 diabetes: a community-based, randomized, controlled trial. *Can J Diabetes* 2021 Oct;45(7):594-600. [doi: [10.1016/j.jcjd.2020.11.012](#)] [Medline: [33582039](#)]
10. Thom D, Willard-Grace R, Hessler D, DeVore D, Prado C, Bodenheimer T, et al. The impact of health coaching on medication adherence in patients with poorly controlled diabetes, hypertension, and/or hyperlipidemia: a randomized controlled trial. *J Am Board Fam Med* 2015;28(1):38-45 [[FREE Full text](#)] [doi: [10.3122/jabfm.2015.01.140123](#)] [Medline: [25567821](#)]

11. Melko CN, Terry PE, Camp K, Xi M, Healey ML. Diabetes health coaching improves medication adherence: a pilot study. *Am J Lifestyle Med* 2009 Oct 22;4(2):187-194. [doi: [10.1177/1559827609351131](https://doi.org/10.1177/1559827609351131)]
12. Coronato A, Naeem M, De Pietro G, Paragliola G. Reinforcement learning for intelligent healthcare applications: a survey. *Artif Intell Med* 2020 Sep;109:101964. [doi: [10.1016/j.artmed.2020.101964](https://doi.org/10.1016/j.artmed.2020.101964)] [Medline: [34756216](https://pubmed.ncbi.nlm.nih.gov/34756216/)]
13. Moodie EEM, Chakraborty B, Kramer MS. Q-learning for estimating optimal dynamic treatment rules from observational data. *Can J Stat* 2012 Dec 01;40(4):629-645 [FREE Full text] [doi: [10.1002/cjs.11162](https://doi.org/10.1002/cjs.11162)] [Medline: [23355757](https://pubmed.ncbi.nlm.nih.gov/23355757/)]
14. Chakraborty B, Moodie EEM. *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*. New York, NY: Springer; 2013.
15. Tejedor M, Woldaregay A, Godtliebsen F. Reinforcement learning application in diabetes blood glucose control: a systematic review. *Artif Intell Med* 2020 Apr;104:101836. [doi: [10.1016/j.artmed.2020.101836](https://doi.org/10.1016/j.artmed.2020.101836)] [Medline: [32499004](https://pubmed.ncbi.nlm.nih.gov/32499004/)]
16. Javad MOM, Agboola S, Jethwani K, Zeid I, Kamarthi S. Reinforcement learning algorithm for blood glucose control in diabetic patients. 2015 Presented at: ASME 2015 International Mechanical Engineering Congress and Exposition. Volume 14: Emerging Technologies; Safety Engineering and Risk Analysis; Materials: Genetics to Structures; November 13-19, 2015; Houston, TX. [doi: [10.1115/imece2015-53420](https://doi.org/10.1115/imece2015-53420)]
17. Vrabie D, Vamvoudakis KG, Lewis FL. *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*. London, UK: Institution of Engineering and Technology; 2012.
18. Ngo PD, Wei S, Holubova A, Muzik J, Godtliebsen F. Reinforcement-learning optimal control for type-1 diabetes. 2018 Presented at: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics; March 4-7, 2018; Las Vegas, NV. [doi: [10.1109/bhi.2018.8333436](https://doi.org/10.1109/bhi.2018.8333436)]
19. Ngo PD, Wei S, Holubová A, Muzik J, Godtliebsen F. Control of blood glucose for type-1 diabetes by using reinforcement learning with feedforward algorithm. *Comput Math Methods Med* 2018;2018:4091497. [doi: [10.1155/2018/4091497](https://doi.org/10.1155/2018/4091497)] [Medline: [30693047](https://pubmed.ncbi.nlm.nih.gov/30693047/)]
20. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. *J Med Internet Res* 2017 Oct 10;19(10):e338 [FREE Full text] [doi: [10.2196/jmir.7994](https://doi.org/10.2196/jmir.7994)] [Medline: [29017988](https://pubmed.ncbi.nlm.nih.gov/29017988/)]
21. Sun X, Bee Y, Lam SW, Liu Z, Zhao W, Chia SY, et al. Effective treatment recommendations for type 2 diabetes management using reinforcement learning: treatment recommendation model development and validation. *J Med Internet Res* 2021 Jul 22;23(7):e27858 [FREE Full text] [doi: [10.2196/27858](https://doi.org/10.2196/27858)] [Medline: [34292166](https://pubmed.ncbi.nlm.nih.gov/34292166/)]
22. Lauffenburger JC, Yom-Tov E, Keller PA, McDonnell ME, Bessette LG, Fontanet CP, et al. REinforcement learning to improve non-adherence for diabetes treatments by Optimising Response and Customising Engagement (REINFORCE): study protocol of a pragmatic randomised trial. *BMJ Open* 2021 Dec 03;11(12):e052091 [FREE Full text] [doi: [10.1136/bmjopen-2021-052091](https://doi.org/10.1136/bmjopen-2021-052091)] [Medline: [34862289](https://pubmed.ncbi.nlm.nih.gov/34862289/)]
23. Sherifali D, Brozic A, Agema P, Gerstein H, Punthakee Z, McInnes N, et al. The diabetes health coaching randomized controlled trial: rationale, design and baseline characteristics of adults living with type 2 diabetes. *Can J Diabetes* 2019 Oct;43(7):477-482. [doi: [10.1016/j.cjcd.2018.10.004](https://doi.org/10.1016/j.cjcd.2018.10.004)] [Medline: [30578164](https://pubmed.ncbi.nlm.nih.gov/30578164/)]
24. Bradley C, Todd C, Gorton T, Symonds E, Martin A, Plowright R. The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. *Qual Life Res* 1999;8(1-2):79-91. [doi: [10.1023/a:1026485130100](https://doi.org/10.1023/a:1026485130100)] [Medline: [10457741](https://pubmed.ncbi.nlm.nih.gov/10457741/)]
25. Toobert D, Hampson S, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. *Diabetes Care* 2000 Jul;23(7):943-950. [doi: [10.2337/diacare.23.7.943](https://doi.org/10.2337/diacare.23.7.943)] [Medline: [10895844](https://pubmed.ncbi.nlm.nih.gov/10895844/)]
26. 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](https://doi.org/10.1016/0168-8510(90)90421-9)] [Medline: [10109801](https://pubmed.ncbi.nlm.nih.gov/10109801/)]
27. Optimizing health coaching for patients with type 2 diabetes using machine learning: a pilot study. GitHub. URL: [https://github.com/hamilton-health-sciences/diabetes\\_coaching](https://github.com/hamilton-health-sciences/diabetes_coaching) [accessed 2022-09-05]
28. Evaluating the Effect of a Diabetes Health Coach in Individuals With Type 2 Diabetes (DiabCoach). ClinicalTrials.gov. URL: <https://clinicaltrials.gov/ct2/show/NCT02128815> [accessed 2022-09-05]
29. Tang Y, Agrawal S. Discretizing continuous action space for on-policy optimization. 2020 Presented at: The Thirty-Fourth AAAI Conference on Artificial Intelligence; New York, NY; February 7-12, 2020 p. 5981-5988. [doi: [10.1609/aaai.v34i04.6059](https://doi.org/10.1609/aaai.v34i04.6059)]
30. Guryanov A. Histogram-based algorithm for building gradient boosting ensembles of piecewise linear decision trees. In: van der Aalst WMP, Batagelj V, Ignatov DI, Khachay M, Kuskova V, Kutuzov A, et al, editors. *Analysis of Images, Social Networks and Texts 8th International Conference, AIST 2019, Kazan, Russia, July 17-19, 2019, Revised Selected Papers*. Cham: Springer; 2019:39-50.
31. Hastie T, Friedman J, Tibshirani R. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer; 2001.
32. Chan J, Lim LL, Wareham NJ, Shaw JE, Orchard TJ, Zhang P, et al. The Lancet Commission on diabetes: using data to transform diabetes care and patient lives. *Lancet* 2021 Dec 19;396(10267):2019-2082. [doi: [10.1016/S0140-6736\(20\)32374-6](https://doi.org/10.1016/S0140-6736(20)32374-6)] [Medline: [33189186](https://pubmed.ncbi.nlm.nih.gov/33189186/)]

33. O'Reilly DJ, Blackhouse G, Bowen J, Brozic A, Agema P, Punthakee Z, et al. Economic analysis of a diabetes health coaching intervention for adults living with type 2 diabetes: a single-centre evaluation from a community-based randomized controlled trial. *Can J Diabetes* 2022 Mar;46(2):165-170. [doi: [10.1016/j.jcjd.2021.08.003](https://doi.org/10.1016/j.jcjd.2021.08.003)] [Medline: [35115250](https://pubmed.ncbi.nlm.nih.gov/35115250/)]
34. Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. *Can J Cardiol* 2022 Feb;38(2):204-213 [[FREE Full text](#)] [doi: [10.1016/j.cjca.2021.09.004](https://doi.org/10.1016/j.cjca.2021.09.004)] [Medline: [34534619](https://pubmed.ncbi.nlm.nih.gov/34534619/)]

## Abbreviations

**ADDQoL:** Audit of Diabetes-Dependent Quality of Life

**DSCA:** Diabetes Self-Care Activities

**HbA<sub>1c</sub>:** hemoglobin A<sub>1c</sub>

**LOOCV:** leave-one-out cross-validation

**QoL:** quality of life

**T2D:** type 2 diabetes

*Edited by A Mavragani; submitted 21.03.22; peer-reviewed by I Donadello; comments to author 13.04.22; revised version received 06.06.22; accepted 07.09.22; published 13.09.22.*

*Please cite as:*

*Di S, Petch J, Gerstein HC, Zhu R, Sherifali D*

*Optimizing Health Coaching for Patients With Type 2 Diabetes Using Machine Learning: Model Development and Validation Study*

*JMIR Form Res* 2022;6(9):e37838

URL: <https://formative.jmir.org/2022/9/e37838>

doi: [10.2196/37838](https://doi.org/10.2196/37838)

PMID: [36099006](https://pubmed.ncbi.nlm.nih.gov/36099006/)

©Shuang Di, Jeremy Petch, Hertz C Gerstein, Ruoqing Zhu, Diana Sherifali. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 13.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Postoperative Outcomes of a Digital Rehabilitation Program After Total Knee Arthroplasty: Retrospective, Observational Feasibility Study

Mindy Hong<sup>1\*</sup>, MSPH, PhD; Joey Loeb<sup>1\*</sup>, DPT; Manshu Yang<sup>2\*</sup>, PhD; Jeannie F Bailey<sup>3\*</sup>, PhD

<sup>1</sup>Hinge Health, Inc, San Francisco, CA, United States

<sup>2</sup>Department of Psychology, University of Rhode Island, Kingston, RI, United States

<sup>3</sup>Department of Orthopaedic Surgery, University of California San Francisco, San Francisco, CA, United States

\* all authors contributed equally

**Corresponding Author:**

Mindy Hong, MSPH, PhD

Hinge Health, Inc.

455 Market Street

Suite 700

San Francisco, CA, 94105

United States

Phone: 1 615 498 8308

Email: [mindy.hong@hingehealth.com](mailto:mindy.hong@hingehealth.com)

## Abstract

**Background:** Surgery can sometimes be the best solution for chronic musculoskeletal pain, but presurgical preparation and postsurgical rehabilitation are often required to achieve the maximum benefits. A digital musculoskeletal surgical care program was developed to support the population of patients undergoing total knee arthroplasty.

**Objective:** We aimed to demonstrate safety, engagement, and acceptability and explore clinical outcomes, health care use, and satisfaction among participants of a digital musculoskeletal surgical care program who were undergoing total knee arthroplasty.

**Methods:** A retrospective, observational feasibility study comparing digital musculoskeletal surgical care program participants to a comparison group was conducted. The intervention group registered for a digital musculoskeletal surgical care program, which included health coaches, physical therapists, and tailored exercises and educational articles to provide preoperative and postoperative support to patients who had recently undergone total knee arthroplasty. Comparison group members received standard-of-care treatment. Engagement (number of exercise therapy sessions and educational articles accessed per week) and acceptability (Net Promoter Score) were examined among intervention group participants. Descriptive statistics for postoperative outcomes, including safety (postoperative complications), clinical improvement (pain, function, anxiety, and depression), and health care use and experiences (length of hospital stay, surgery satisfaction, and physical therapy adherence), were reported for both groups. Differences among postoperative results were compared by using the independent samples 2-tailed *t* test or Mann-Whitney test for continuous outcomes and the Fisher exact test or chi-square test for categorical outcomes.

**Results:** Of the 53 participants (intervention group: *n*=22; comparison group: *n*=31) who were included in this study, 35 (66%) were female and 25 (47%) were aged from 45 to 60 years. On average, the intervention group completed 23 exercise sessions, read 2.7 educational articles, sent 45.5 texts to their health coaches, and were actively engaged for 6 weeks after their operation. Among 21 participants, 14 (67%) self-reported as *promoters* on the Net Promoter Score scale. Intervention group members reported fewer postoperative complications (6/22, 27%) than the comparison group (15/31, 48%), and they experienced better outcomes with regard to function (Knee Injury and Osteoarthritis Outcome Score—Physical Function Short Form—intervention group: mean 23.0; comparison group: mean 32.5), depression (Patient Health Questionnaire 2-Item—intervention group: mean 0.4; comparison group: mean 1.6), anxiety (General Anxiety Disorder 2-Item—intervention group: mean 0.6; comparison group: mean 1.5), and impressions of change (Patient Global Impression of Change—intervention group: median 7.0; comparison group: median 6.0). Intervention group participants also reported less health care use, better adherence to their physical therapy exercises, and higher surgery satisfaction.

**Conclusions:** Our digital musculoskeletal surgical care program shows promising levels of engagement and acceptability among those who recently underwent total knee arthroplasty. The surgical care program may also help with improving postsurgical complications and clinical outcomes and lowering health care use.

(JMIR Form Res 2022;6(9):e40703) doi:[10.2196/40703](https://doi.org/10.2196/40703)

## KEYWORDS

total knee arthroplasty; surgical; digital intervention; musculoskeletal; telemedicine

## Introduction

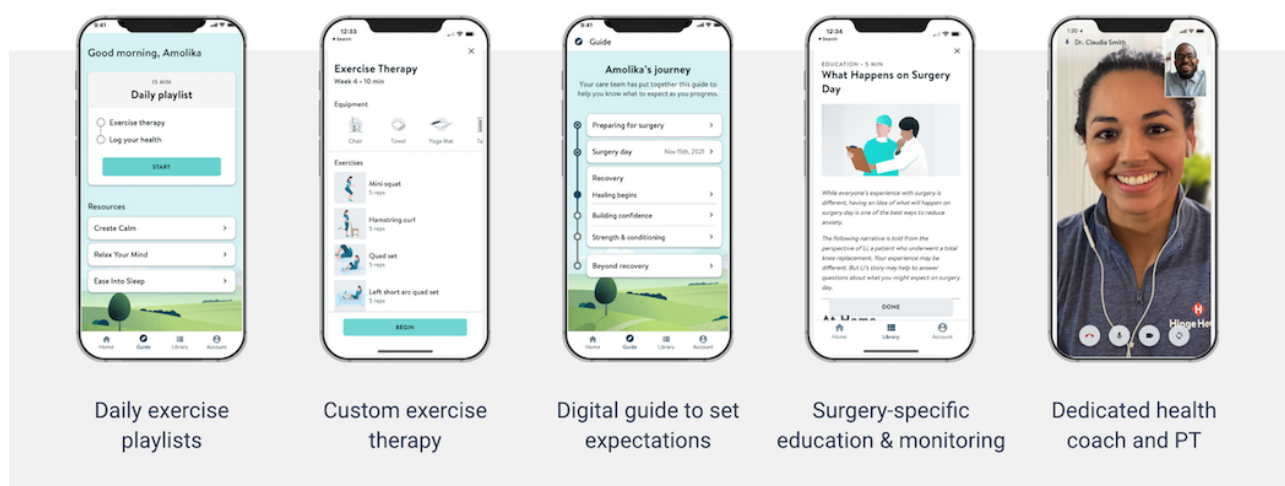
Chronic musculoskeletal pain is a leading cause of disability and increased health care costs in the United States, affecting over 50 million people and resulting in an estimated total lost productivity cost of US \$296 million per year [1]. Evidence-based clinical guidelines typically recommend performing nonsurgical interventions before performing invasive procedures for people with chronic musculoskeletal pain [2]. However, these treatments are not always effective for everyone, and when these treatments are inadequate, surgery can be recommended [3]. For example, total knee arthroplasties have proven to be successful in improving pain, mobility, and quality of life for many with chronic knee pain [4,5]. However, despite substantial improvements in surgical techniques and treatments, approximately 20% of patients report dissatisfaction following total knee arthroplasty [5,6]. The causes of dissatisfaction appear to be multifactorial; unmet expectations, the inability to engage in postoperative rehabilitation, and limited pain relief can affect satisfaction after total knee arthroplasty [6,7].

For surgery outcomes to be successful, it is important for patients to adhere to both preoperative rehabilitation and postoperative rehabilitation [5,8-11]. However, almost half of adults do not engage in and adhere to postoperative rehabilitation [12]. This is due to gaps in social support, poorly managed expectations, and a lack of education regarding the benefits of postoperative rehabilitation [7,13]. Furthermore, despite the benefits of receiving social support after total knee arthroplasty [14], not all patients have the social support they need.

In order to address care gaps and their subsequent impacts on both care quality and clinical outcomes, we developed a digital surgical care program consisting of medical, social, and educational support (Figure 1). The digital musculoskeletal surgical care program's goal was to support and help patients throughout the preoperative and postoperative stages of surgery. The program included surgical health coaches and certified physical therapists who worked with participants to help them prepare for and recover from surgery, answer questions, and customize participants' plans of care. In addition, the program also included customized exercise therapy sessions that followed the protocols created by participants' surgeons and aimed to strengthen and rehabilitate participants. The participants also received tailored education articles on lifestyle management and recovery tips that helped prepare them for the preoperative and postoperative phases of surgery. All materials and services were accessed through the digital musculoskeletal surgical care program app.

The preoperative phase of the surgical care program started up to 8 weeks prior to surgery. Participants worked with health coaches to achieve goals that supported the best surgical outcomes, such as reducing presurgery anxiety or creating healthier eating habits for recovery. Participants also worked with physical therapists to strengthen muscles that supported their new joints to allow for optimal recovery. The postoperative phase lasted for 12 weeks after surgery. During this time, participants received support and accountability training for achieving their recovery goals, as per their rehabilitation plans.

**Figure 1.** The digital musculoskeletal surgical care program provides users with a tailored program containing medical, social, and educational support. PT: physical therapist.





The program was complemented by the plans that were recommended and directed by participants' surgeons. As such, participants still attended their recommended in-person physical therapy sessions. Participants informed their care program health coaches of their in-person physical therapy exercises, and information on these exercises was loaded into the app to assist with tracking progress.

The goal of this study was to determine the feasibility of our novel surgical care program. The primary objective was to demonstrate safety, engagement, and acceptability among the participants of the surgical care program. The secondary objectives included exploring postoperative clinical improvement and comparing the health care use and experiences of surgical care program participants against those of a comparison group.

Patients who need orthopedic surgery typically benefit from the social support provided through caregivers or family members, but not everyone is able to receive assistance. Therefore, our digital surgical care program aims to provide the needed social support and improve care quality and clinical outcomes both before and after the surgery process. The results from this study will contribute to developing larger-scale studies that will provide evidence that this digital surgical care program effectively addresses care gaps.

## Methods

### Study Design

A retrospective, observational feasibility study comparing digital musculoskeletal surgical care program participants (herein, the *intervention group*) to a comparison group was conducted.

### Study Participants

Participants who were eligible for the intervention were identified based on our inclusion and exclusion criteria and the information they provided in the app. The inclusion criteria were patients who had a smartphone, patients who created an account for the digital musculoskeletal surgical care program, patients who completed the app questionnaire at least 4 weeks prior to starting the surgical care program, patients who enrolled in the knee surgical care program, patients who underwent knee replacement surgery 6 or more weeks before survey data collection, and English-speaking patients.

Comparison group participants were recruited through a proprietary research panel maintained by Momentive. The panel includes over 100 million people who are invited to take part in surveys. We screened panelists by age and included those who had not previously participated in Hinge Health programs and underwent knee replacement surgery between June and December 2021.

### Ethics Approval

This study (reference number: 20160949) was reviewed and approved by the WIRB-Copernicus Group Institutional Review Board (Office for Human Research Protections and Food and Drug Administration Institutional Review Board registration number: IRB00000533). Intervention group participants acknowledged and provided research consent. The institutional

review board deemed the comparison group participants exempt from providing informed consent.

### Variables

#### *Overview of Outcomes*

For the primary objective of this feasibility study, we examined engagement, acceptability, and safety outcomes to demonstrate that it was possible to implement the program and this study. We also explored proximal and distal outcomes that we hypothesized would be influenced by the program. These included clinical outcomes (pain, function, depression, and anxiety), health care use (length of hospital stay and physical therapy adherence), and health care experiences (surgery satisfaction). This study collected data on outcomes regarding engagement and acceptability from the intervention group. Data on safety, clinical improvement, and health care use and experience outcomes were collected from both the intervention group and the comparison group.

#### *Engagement and Acceptability*

To assess the intervention group's engagement, data on the number of app-based exercise therapy sessions and educational articles accessed per week were collected through the app. Acceptability was measured through the Net Promoter Scores (NPSs) that participants provided for the following question: "On a scale of 0-10, how likely is it that you would recommend the surgical program to a friend or colleague" (0=not likely; 10=very likely)? NPSs ranging from 0 to 6 were labeled as *detractors*, NPSs of 7 to 8 were labeled as *passives*, and NPSs of 9 to 10 were labeled as *promoters*.

#### *Safety*

Safety was based on self-reported postsurgical complications, which included wound infections requiring antibiotics or surgery, blood clots or deep vein thrombosis, stiffness requiring manipulation under anesthesia, falls, surgery on the same knee for other reasons, and general soreness.

#### *Clinical Improvement*

The clinical outcomes included pain improvement, function, anxiety, and depression. Patient Global Impression of Change (PGIC) scores were measured for both groups through the following question: "Compared to before the surgery, how would you rate your knee pain now" (1=much worse; 7=much improved)? Function was measured through the Knee Injury and Osteoarthritis Outcome Score-Physical Function Short Form (KOOS-PS; 0=no difficulty; 100=extreme difficulty). Anxiety and depression were measured through the General Anxiety Disorder 2-Item (GAD-2) and Patient Health Questionnaire 2-Item (PHQ-2), respectively.

#### *Health Care Use and Experiences*

We measured the length of stay in the hospital after surgery based on the following question: "How many nights did you spend in the hospital or surgery facility after your surgery?" Surgery satisfaction was measured based on the following question: "Overall, how satisfied were you with the results of your knee replacement surgery" (0=very dissatisfied; 3=very satisfied)? Physical therapy adherence was measured through

the following question: “How often did you perform home exercise sessions as recommended by your in-person physical therapist” (0=never; 5=always)?

### Data Sources

The web-based program registration process provided baseline demographic data for the intervention group. Surveys that evaluated postoperative outcomes were emailed to both the intervention group and the comparison group. Intervention group respondents received gift cards worth US \$30 upon the completion of the survey.

### Statistical Methods

Because this was a feasibility study, no formal sample size calculations were conducted. Summary statistics were estimated for the baseline demographic characteristics of the intervention group. Descriptive statistics for the postoperative outcomes of both the intervention group and the comparison group were reported, including means with SDs and medians with IQRs.

Differences between the postoperative results of the two groups were compared by using the independent samples 2-tailed *t* test or Mann-Whitney *U* test for continuous or ordinal outcomes. Categorical variables were analyzed by using the Fisher exact test or chi-square test. Analyses were performed in R version 4.0.5 (R Foundation for Statistical Computing).

## Results

### Sample Characteristics

Altogether, 53 participants completed the study (intervention group: *n*=22; comparison group: *n*=31). Table 1 shows the demographics and the months of surgery for both groups. Apart from age (intervention group participants were significantly older than comparison group participants; *P*<.001), no differences between the two groups were detected at baseline. The majority of the sample was aged from 45 to 60 years (25/53, 47%), was female (35/53, 66%), and received surgery in September (11/53, 21%) and October (11/53, 21%).

**Table 1.** Study participant characteristics.

Characteristics	Intervention group ( <i>n</i> =22), <i>n</i> (%)	Comparison group ( <i>n</i> =31), <i>n</i> (%)	All participants ( <i>N</i> =53), <i>n</i> (%)	<i>P</i> value
<b>Age (years)</b>				<.001 <sup>a</sup>
30-44	0 (0)	4 (13)	4 (8)	
45-60	5 (23)	20 (65)	25 (47)	
>60	17 (77)	7 (23)	24 (45)	
<b>Sex</b>				.57 <sup>b</sup>
Male	6 (27)	12 (39)	18 (34)	
Female	16 (73)	19 (61)	35 (66)	
<b>Month of surgery<sup>c</sup></b>				.99 <sup>a</sup>
June	1 (5)	2 (7)	3 (6)	
July	3 (14)	5 (16)	8 (15)	
August	2 (9)	4 (13)	6 (11)	
September	4 (18)	7 (23)	11 (21)	
October	5 (23)	6 (19)	11 (21)	
November	4 (18)	3 (10)	7 (13)	
December	3 (14)	4 (13)	7 (13)	

<sup>a</sup>Fisher exact test.

<sup>b</sup>Chi-square test with continuity correction.

<sup>c</sup>Months of surgery are from 2021.

### Engagement and Acceptability (Intervention Only)

On average, intervention group participants completed 23 (SD 34.8) exercise therapy sessions, read 2.7 (SD 5.3) educational articles, sent 45.5 (SD 51.7) text messages to their health coaches, and were actively engaged for 6 (SD 6.7) weeks after their operation. Among 21 program participants, 14 (67%) were promoters on the NPS.

### Safety

The percentage of participants who reported postsurgical complications was higher in the comparison group versus the

intervention group by 21%. The reported complications among comparison group members were stiffness (4/15, 27%), surgery on the same knee for other reasons (4/15, 27%), blood clots or deep vein thrombosis (4/15, 27%), wound infections requiring antibiotics (2/15, 13%), general soreness (2/15, 13%), infections requiring surgery (1/15, 7%), and falls (1/15, 7%). The complications among the intervention group were stiffness (4/6, 67%) and vasculitis (ie, a latex allergy; 1/6, 17%). Manipulation for addressing stiffness was recommended for a patient, but the treatment was deferred (1/6, 17%).

## Clinical Improvement

The intervention group reported better knee pain after their surgery compared to that reported by the comparison group (PGIC score: mean 7.0 vs mean 6.0;  $P=.06$ ). The intervention group reported better function scores (KOOS-PS: mean 23.0 vs mean 32.5;  $P=.049$ ) than those reported by the comparison group. Intervention group members also reported lower anxiety (GAD-2 score: mean 0.6 vs mean 1.5;  $P=.01$ ) and depression (PHQ-2 score: mean 0.4 vs mean 1.6;  $P=.004$ ) scores than those reported by the comparison group.

## Health Care Use and Experiences

The median length of hospital stay was 1 (IQR 0) night for the intervention group and 1 (IQR 1) night for the comparison group. There was a significant difference ( $W=275$ ;  $P=.009$ ) in the lengths of hospital stay between the intervention group and the comparison group. Additionally, the intervention group showed better adherence to the exercises that were recommended by their physical therapists ( $P=.06$ ) and reported higher satisfaction with their surgery experience ( $P=.06$ ). Table 2 shows the postoperative outcomes of the intervention and comparison groups.

**Table 2.** Postoperative outcomes.

Outcomes	Intervention group (n=22)	Comparison group (n=31)	All participants (N=53)	P value
<b>Safety (postsurgical complications), n (%)</b>				
No complications	16 (73)	16 (52)	32 (60)	.12 <sup>a</sup>
Complications	6 (27)	15 (48)	21 (40)	
<b>Clinical outcomes</b>				
Patient Global Impression of Change score, median (IQR)	7.0 (1.0)	6.0 (2.0)	7.0 (1.0)	.06 <sup>b</sup>
Knee Injury and Osteoarthritis Outcome Score–Physical Function Short Form, mean (SD)	23.0 (17.4)	32.5 (15.9)	28.6 (17.0)	.049 <sup>c</sup>
General Anxiety Disorder 2-Item score, mean (SD)	0.6 (1.3)	1.5 (1.6)	1.1 (1.5)	.01 <sup>b</sup>
Patient Health Questionnaire 2-Item score, mean (SD)	0.4 (1.1)	1.6 (1.8)	1.1 (1.7)	.004 <sup>b</sup>
<b>Health care use and experiences</b>				
Length of hospital stay (number of nights), median (IQR)	1.0 (0)	1.0 (1.0)	1.0 (0)	.009 <sup>b</sup>
Physical therapy adherence score, median (IQR)	4.0 (2.0)	3.0 (2.5)	3.0 (3.0)	.06 <sup>b</sup>
Surgery satisfaction score, median (IQR)	3.0 (1.0)	2.0 (1.0)	2.0 (1.0)	.06 <sup>b</sup>

<sup>a</sup>Chi-square test.

<sup>b</sup>Mann-Whitney *U* test.

<sup>c</sup>Independent samples *t* test.

## Discussion

### Principal Results

This study aimed to assess the feasibility of a novel, digital surgical care program for total knee arthroplasty by (1) examining patient engagement, acceptability, and safety and (2) exploring clinical improvements as well as health care use and experiences. First, we posit that a program that offers clinical, educational, and social support is viable and that it would be well received by persons planning to undergo a total knee arthroplasty. We found that on average, the intervention group remained active in the program by engaging in 23 exercise sessions, reading 2.7 educational articles, and sending 45.5 text messages to their coaches. Satisfaction among the intervention group was high, as two-thirds of participants (14/21, 67%) were NPS promoters.

We also demonstrated that the intervention was safe. The most frequently reported complication from the intervention group was stiffness, which can be expected postoperatively. Manipulation under anesthesia for stiffness was recommended

for a patient by the surgeon but was deferred by the patient. The remaining complication (a latex allergy) was unrelated to the surgery itself. The intervention group also experienced fewer adverse events and complications than those experienced by the comparison group. One possible explanation is that the intervention helped participants adhere to postsurgical exercise regimens and avoid complications. Another explanation is the presence of unmeasured confounding variables. For example, intervention group members may have received a higher quality of in-person care compared to the care that the comparison group received.

The results of this study will be used in two formative projects. In the first project, we will use the lessons learned in this study to refine and improve the program. For example, we learned more about the nature of the complications experienced by both the intervention group and the comparison group. We may prepare additional tools and resources for participants about how to identify complications, when to contact providers for help, and how to best recover from complications. In the second project, we will use the estimates from this preliminary study

to develop a larger observational study that is powered to detect statistically significant differences between an intervention group and comparison group. Furthermore, implementing the surgical care program at scale will allow us to further evaluate intervention effectiveness in a more generalizable setting.

### Comparison With Prior Work

We provide preliminary evidence that an intervention that offers clinical and social support is associated with better clinical outcomes, including better impressions of change, function, anxiety, and depression scores after surgery. These results are consistent with those of a qualitative study that found that patients who had a positive total knee arthroplasty experience often reported having adequate social and clinical support [15]. These results are also consistent with those of studies reporting that social support has a positive impact on pain, function, mental health, and patient satisfaction among patients who have undergone joint replacement surgery [14-16]. A meta-analysis found that the presence of social support had a beneficial effect on total Western Ontario and McMaster Universities Arthritis Index scores (mean difference: 2.88, 95% CI 1.30-4.46), which are used to measure pain, functional limitations, and stiffness. The same meta-analysis also found social support to be positively associated with improvements in knee pain and function (total Oxford Knee Scores—mean difference: 0.29, 95% CI 0.12-0.45) [14]. We speculate that improvements in clinical outcomes could be the result of having a care team that consists of health coaches, physical therapists, and physicians who support patients and offer medical advice throughout recovery. Our study also explored health care use and experiences and showed shorter lengths of hospital stay, better physical therapy adherence, and higher levels of satisfaction in the intervention group. Similarly, studies have reported that inadequate social support and unmet expectations result in higher rates of dissatisfaction, poorer adherence to outpatient therapy, and poorer outcomes [5-7,10]. In a prospective study of 1703 participants that examined satisfaction after total knee arthroplasty, unmet expectations was one of the strongest predictors of dissatisfaction in a comparison between dissatisfied and satisfied participants (49% vs 6%) [5].

### Strengths and Limitations

The strengths of this study include evaluating multiple outcomes to demonstrate program feasibility, including engagement, acceptability, safety, clinical, and health care use and experience outcomes. This study also included a comparison group for comparing postoperative outcomes.

This study however also presents limitations. As with most retrospective studies, there is a risk of bias associated with data that are collected prior to the start of a study. This study compared the postoperative outcomes of the intervention and comparison groups. A future, larger-scale observational study that compares both preoperative outcomes and postoperative outcomes with those of a comparison group, as well as surgeries in other pathways (eg, total hip arthroplasty), can provide more insight into the effectiveness of a digital musculoskeletal surgical care program. Lastly, this study was designed to demonstrate feasibility with a small set of early surgical care program members and a convenience sample comparison group. As such, there are differences between the two groups that may have influenced our exploratory clinical outcomes. The preliminary findings of this study are encouraging, and they will be used as the basis for the next steps in program development, that is, in research where we will adjust for potential confounding factors.

### Conclusions

Our digital musculoskeletal surgical care program, which provides social support, medical advice, and education to those who have recently undergone total knee arthroplasty, is feasible and acceptable. We demonstrated engagement, satisfaction, and safety among the participants of the program. Additionally, compared to the comparison group, this study showed preliminary evidence of improved clinical outcomes, lower health care use, and higher satisfaction among intervention group participants. Based on the results of this feasibility study, a larger-scale observational study can build on the findings of this study to further evaluate the effectiveness of the surgical care program.

---

### Acknowledgments

Hinge Health, Inc, provided the digital musculoskeletal surgical care program to participants.

---

### Authors' Contributions

MH, MY and JFB designed the study. MH and MY analyzed the data. MH, JL, MY, and JFB interpreted the data and were major contributors in writing the manuscript. MH, JL, MY, and JFB read and approved the final manuscript.

---

### Conflicts of Interest

MH and JL are Hinge Health, Inc, employees and have equity interest in Hinge Health, Inc.

---

### References

1. Yong RJ, Mullins PM, Bhattacharyya N. Prevalence of chronic pain among adults in the United States. *Pain* 2022 Feb 01;163(2):e328-e332. [doi: [10.1097/j.pain.0000000000002291](https://doi.org/10.1097/j.pain.0000000000002291)] [Medline: [33990113](https://pubmed.ncbi.nlm.nih.gov/33990113/)]

2. National Institute for Health and Care Excellence (NICE). Chronic Pain (Primary and Secondary) in Over 16s: Assessment of All Chronic Pain and Management of Chronic Primary Pain. London, United Kingdom: National Institute for Health and Care Excellence (NICE); Apr 07, 2021.
3. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis Cartilage* 2008 Feb;16(2):137-162 [FREE Full text] [doi: [10.1016/j.joca.2007.12.013](https://doi.org/10.1016/j.joca.2007.12.013)] [Medline: [18279766](https://pubmed.ncbi.nlm.nih.gov/18279766/)]
4. Miettinen HJA, Mäkirinne-Kallio N, Kröger H, Miettinen SSA. Health-related quality of life after hip and knee arthroplasty operations. *Scand J Surg* 2021 Sep;110(3):427-433 [FREE Full text] [doi: [10.1177/1457496920952232](https://doi.org/10.1177/1457496920952232)] [Medline: [32862793](https://pubmed.ncbi.nlm.nih.gov/32862793/)]
5. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? *Clin Orthop Relat Res* 2010 Jan;468(1):57-63 [FREE Full text] [doi: [10.1007/s11999-009-1119-9](https://doi.org/10.1007/s11999-009-1119-9)] [Medline: [19844772](https://pubmed.ncbi.nlm.nih.gov/19844772/)]
6. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient dissatisfaction following total knee arthroplasty: A systematic review of the literature. *J Arthroplasty* 2017 Dec;32(12):3854-3860. [doi: [10.1016/j.arth.2017.07.021](https://doi.org/10.1016/j.arth.2017.07.021)] [Medline: [28844632](https://pubmed.ncbi.nlm.nih.gov/28844632/)]
7. Bakaa N, Chen LH, Carlesso L, Richardson J, Shanthanna H, Macedo L. Understanding barriers and facilitators of exercise adherence after total-knee arthroplasty. *Disabil Rehabil*. Epub ahead of print 2021 Aug 16. [doi: [10.1080/09638288.2021.1965232](https://doi.org/10.1080/09638288.2021.1965232)] [Medline: [34397309](https://pubmed.ncbi.nlm.nih.gov/34397309/)]
8. Longo UG, Matarese M, Arcangeli V, Alciati V, Candela V, Facchinetti G, et al. Family caregiver strain and challenges when caring for orthopedic patients: A systematic review. *J Clin Med* 2020 May 16;9(5):1497 [FREE Full text] [doi: [10.3390/jcm9051497](https://doi.org/10.3390/jcm9051497)] [Medline: [32429398](https://pubmed.ncbi.nlm.nih.gov/32429398/)]
9. Pozzi F, Snyder-Mackler L, Zeni J. Physical exercise after knee arthroplasty: a systematic review of controlled trials. *Eur J Phys Rehabil Med* 2013 Dec;49(6):877-892 [FREE Full text] [Medline: [24172642](https://pubmed.ncbi.nlm.nih.gov/24172642/)]
10. Gilbey HJ, Ackland TR, Wang AW, Morton AR, Trouchet T, Tapper J. Exercise improves early functional recovery after total hip arthroplasty. *Clin Orthop Relat Res* 2003 Mar(408):193-200. [doi: [10.1097/00003086-200303000-00025](https://doi.org/10.1097/00003086-200303000-00025)] [Medline: [12616059](https://pubmed.ncbi.nlm.nih.gov/12616059/)]
11. Pisters MF, Veenhof C, Schellevis FG, Twisk JWR, Dekker J, De Bakker DH. Exercise adherence improving long-term patient outcome in patients with osteoarthritis of the hip and/or knee. *Arthritis Care Res (Hoboken)* 2010 Aug;62(8):1087-1094 [FREE Full text] [doi: [10.1002/acr.20182](https://doi.org/10.1002/acr.20182)] [Medline: [20235201](https://pubmed.ncbi.nlm.nih.gov/20235201/)]
12. Groen JW, Stevens M, Kersten RFMR, Reininga IHF, van den Akker-Scheek I. After total knee arthroplasty, many people are not active enough to maintain their health and fitness: an observational study. *J Physiother* 2012;58(2):113-116 [FREE Full text] [doi: [10.1016/S1836-9553\(12\)70091-7](https://doi.org/10.1016/S1836-9553(12)70091-7)] [Medline: [22613241](https://pubmed.ncbi.nlm.nih.gov/22613241/)]
13. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. *Man Ther* 2010 Jun;15(3):220-228 [FREE Full text] [doi: [10.1016/j.math.2009.12.004](https://doi.org/10.1016/j.math.2009.12.004)] [Medline: [20163979](https://pubmed.ncbi.nlm.nih.gov/20163979/)]
14. Wylde V, Kunutsor SK, Lenguerrand E, Jackson J, Blom AW, Beswick AD. Association of social support with patient-reported outcomes after joint replacement: a systematic review and meta-analysis. *Lancet Rheumatol* 2019 Nov 01;1(3):e174-e186. [doi: [10.1016/s2665-9913\(19\)30050-5](https://doi.org/10.1016/s2665-9913(19)30050-5)]
15. Goldsmith LJ, Suryaprakash N, Randall E, Shum J, MacDonald V, Sawatzky R, et al. The importance of informational, clinical and personal support in patient experience with total knee replacement: a qualitative investigation. *BMC Musculoskelet Disord* 2017 Mar 24;18(1):127 [FREE Full text] [doi: [10.1186/s12891-017-1474-8](https://doi.org/10.1186/s12891-017-1474-8)] [Medline: [28340610](https://pubmed.ncbi.nlm.nih.gov/28340610/)]
16. Escobar A, Quintana JM, Bilbao A, Azkárate J, Güenaga JI, Arenaza JC, et al. Effect of patient characteristics on reported outcomes after total knee replacement. *Rheumatology (Oxford)* 2007 Jan;46(1):112-119. [doi: [10.1093/rheumatology/ke1184](https://doi.org/10.1093/rheumatology/ke1184)] [Medline: [16735451](https://pubmed.ncbi.nlm.nih.gov/16735451/)]

## Abbreviations

**GAD-2:** General Anxiety Disorder 2-Item

**KOOS-PS:** Knee Injury and Osteoarthritis Outcome Score–Physical Function Short Form

**NPS:** Net Promoter Score

**PGIC:** Patient Global Impression of Change

**PHQ-2:** Patient Health Questionnaire 2-Item

*Edited by A Mavragani; submitted 01.07.22; peer-reviewed by AA Seid, S Sarejloo; comments to author 27.07.22; revised version received 10.08.22; accepted 12.08.22; published 19.09.22.*

*Please cite as:*

*Hong M, Loeb J, Yang M, Bailey JF*

*Postoperative Outcomes of a Digital Rehabilitation Program After Total Knee Arthroplasty: Retrospective, Observational Feasibility Study*

*JMIR Form Res 2022;6(9):e40703*

*URL: <https://formative.jmir.org/2022/9/e40703>*

*doi: [10.2196/40703](https://doi.org/10.2196/40703)*

*PMID: [36121690](https://pubmed.ncbi.nlm.nih.gov/36121690/)*

©Mindy Hong, Joey Loeb, Manshu Yang, Jeannie F Bailey. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 19.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# mHealth-Supported Gender- and Culturally Sensitive Weight Loss Intervention for Hispanic Men With Overweight and Obesity: Single-Arm Pilot Study

David O Garcia<sup>1</sup>, PhD; Luis A Valdez<sup>2</sup>, MPH, PhD; Benjamin Aceves<sup>3</sup>, MPH, PhD; Melanie L Bell<sup>4</sup>, PhD; Brooke A Rabe<sup>4</sup>, PhD; Edgar A Villavicencio<sup>1</sup>, MPH; David G Marrero<sup>1</sup>, PhD; Forest Melton<sup>5</sup>, MS; Steven P Hooker<sup>6</sup>, PhD

<sup>1</sup>Department of Health Promotion Sciences, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

<sup>2</sup>Department of Community Health and Prevention, Dornsife School of Public Health, Drexel University, Philadelphia, PA, United States

<sup>3</sup>School of Public Health, College of Health and Human Services, San Diego State University, San Diego, CA, United States

<sup>4</sup>Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

<sup>5</sup>Department of Clinical and Translational Sciences, College of Medicine, University of Arizona, Tucson, AZ, United States

<sup>6</sup>College of Health and Human Services, San Diego State University, San Diego, CA, United States

**Corresponding Author:**

David O Garcia, PhD

Department of Health Promotion Sciences

Mel and Enid Zuckerman College of Public Health

University of Arizona

3950 S. Country Club

Suite 330

Tucson, AZ, 85714

United States

Phone: 1 520 626 4641

Email: [davidogarcia@arizona.edu](mailto:davidogarcia@arizona.edu)

## Abstract

**Background:** Hispanic men have disproportionate rates of overweight and obesity compared with other racial and ethnic subpopulations. However, few weight loss interventions have been developed specifically for this high-risk group. Furthermore, the use of mobile health (mHealth) technologies to support lifestyle behavior changes in weight loss interventions for Hispanic men is largely untested.

**Objective:** This single-arm pilot study examined the feasibility and acceptability of integrating mHealth technology into a 12-week gender- and culturally sensitive weight loss intervention (GCSWLI) for Hispanic men with overweight and obesity.

**Methods:** A total of 18 Hispanic men (mean age 38, SD 10.9 years; mean BMI 34.3, SD 5.5 kg/m<sup>2</sup>; 10/18, 56% Spanish monolingual) received a GCSWLI, including weekly in-person individual sessions, a daily calorie goal, and prescription of ≥225 minutes of moderate-intensity physical activity per week. mHealth technology support included tailored SMS text messaging, behavior self-monitoring support using Fitbit Charge 2, and weight tracking using a Fitbit Aria Wi-Fi Smart Scale. Changes in weight from baseline to 12 weeks were estimated using a paired 2-tailed *t* test. Descriptive analyses characterized the use of Fitbit and smart scales. Semistructured interviews were conducted immediately after intervention to assess the participants' weight loss experiences and perspectives on mHealth technologies.

**Results:** Of 18 participants, 16 (89%) completed the 12-week assessments; the overall attrition rate was 11.1%. The mean weight loss at week 12 was -4.7 kg (95% CI 7.1 to -2.4 kg; *P*<.001). Participants wore the Fitbit 71.58% (962/1344) of the intervention days and logged their body weight using the smart scale (410/1344, 30.51% of the intervention days). Participants identified barriers to the use of the technology, such as lack of technological literacy and unreliable internet access for the smart scale.

**Conclusions:** Although clinically significant weight loss was achieved by integrating mHealth technology into the GCSWLI, adherence to the prescribed use of technology was modest. Addressing barriers to the use of such technologies identified in our work may help to refine an mHealth intervention approach for Hispanic men.

**Trial Registration:** ClinicalTrials.gov NCT02783521; <https://clinicaltrials.gov/ct2/show/NCT02783521>

**KEYWORDS**

Hispanic; mobile health; mHealth; overweight; obesity; weight loss

## Introduction

### Background

Hispanics, one of the largest and fastest growing racial and ethnic subpopulations in the United States, comprise 17% of the total US population and are projected to reach 28% by 2060 [1]. In parallel with population growth, the prevalence of obesity continues to increase. Hispanic men, in particular, have the highest age-adjusted prevalence of obesity (BMI $\geq$ 30 kg/m<sup>2</sup>; 45.7%) when compared with non-Hispanic Black (41.1%) and Hispanic men (44.7%) [2]. Consequently, the incidence of obesity-related diseases such as type 2 diabetes mellitus and nonalcoholic fatty liver disease, which are associated with obesity, is the highest among Hispanic men [3,4]. Thus, it is surprising that despite the growing prevalence of obesity and obesity-related diseases in Hispanic men, weight loss interventions to reduce their body weight are understudied.

Lifestyle interventions have proven to be an effective strategy for implementing chronic disease prevention approaches that are responsive to weight loss [5]. Mobile health (mHealth) technologies, including the use of mobile phones, tablet computers, mobile apps, and wearable devices such as smart watches, are increasingly being used to promote dietary and physical activity behaviors within lifestyle interventions [6,7]. mHealth diminishes access to care barriers for hard-to-reach populations, increases data collection efficiency and accuracy, and provides health care professionals with a means to communicate with participants more frequently in clinical and nonclinical settings [8,9]. Although interventions using mHealth technologies generally have positive results, the methodological standards of evaluated studies are often low [9].

Hispanics are receptive to using mHealth technologies for preventive care services. In particular, SMS text messages have been used as an effective tool for engaging Hispanic participants in research studies [8-11]. A randomized clinical trial by Rosas et al [12,13] demonstrated the effective use of Fitbit devices among Hispanic participants with overweight or obesity, particularly when complemented by a culturally responsive lifestyle intervention. More recently, Rosas [14] found that Hispanic men who chose to attend web-based videoconference sessions lost significantly more weight when compared with a group who chose to watch prerecorded videos on the web. This finding suggests that providing options to accommodate the preferences of Hispanic men for intervention delivery methods can increase engagement in lifestyle interventions [14]. However, the use of mHealth technologies to support lifestyle behavior changes in weight loss interventions for Hispanic men remains largely untested.

### Objectives

Given the rapid growth of the Hispanic population in the United States and the disproportionate burden of preventable chronic diseases faced by this population, effective and sustainable interventions are urgently required. In this context, engaging Hispanic men with mHealth technology that is linguistically and culturally competent, while also meeting their health needs within chronic disease prevention, may be a potential solution. Earlier formative work by our investigative team showed that an mHealth-supported intervention, in combination with face-to-face counseling, may improve engagement and adherence to behavior change recommendations, particularly for physical activity [15-17]. Here, we report the findings from a single-arm pre-post pilot study that examined the feasibility and acceptability of integrating mHealth technology into a 12-week gender- and culturally sensitive weight loss intervention (GCSWLI) targeting Hispanic men with overweight and obesity. We focused on men because they have been shown to be less likely to participate in lifestyle interventions than women [18,19]. In this regard, men, particularly in the Hispanic population, are understudied.

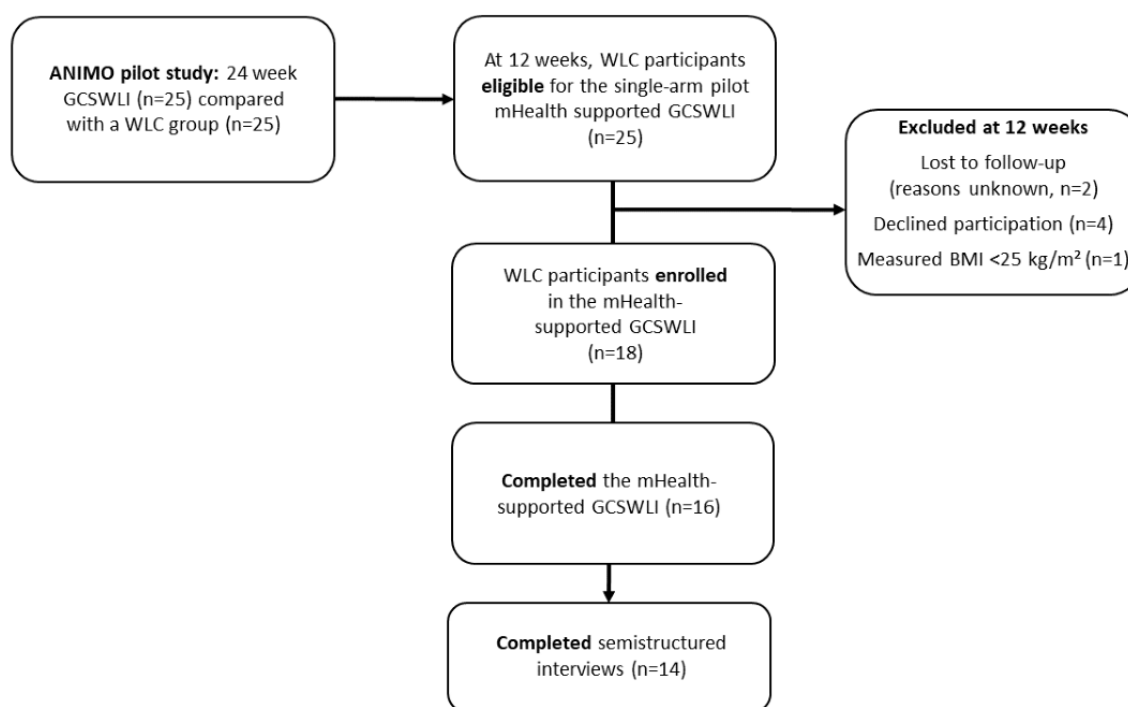
## Methods

### Study Design and Participants

Study participants were part of the ANIMO (a Spanish term for motivation or encouragement) pilot study (ClinicalTrials.gov NCT02783521), a 24-week randomized controlled trial that tested the effects of in-person GCSWLI on body weight in Hispanic males (n=25) compared with a waitlist control (WLC) group (n=25) [20,21]. Full details of the ANIMO study design, protocol for the gender- and culturally supported intervention, and outcome findings for the GCSWLI have been published elsewhere [20,21]. Briefly, 98% of the men were of Mexican-origin descent, 58% reported Spanish as their preferred language at home, and 50% were born outside the United States [20,21]. On completion of the initial 12-week GCSWLI, the WLC participants were eligible to receive the mHealth-supported GCSWLI for 12 weeks (Figure 1). To maintain eligibility, men (1) were to be aged 18 to 64 years; (2) had to have a BMI of 25 to 50 kg/m<sup>2</sup> [22]; (3) had to self-identify as Hispanic; (4) had to have the ability to provide informed consent and complete a health risk assessment before participation; and (5) had to be able to speak, read, and write English or Spanish. Semistructured interviews were conducted after completion of the 12-week intervention to assess participants' weight loss experiences and perspectives on the potential use of mHealth technologies in future intervention efforts.



**Figure 1.** Study flow diagram for the mobile health (mHealth)–supported gender- and culturally sensitive weight loss intervention (GCSWLI). WLC: waitlist control.



### Ethics Approval

All study activities were conducted at the University of Arizona Collaboratory for Metabolic Disease Prevention and Treatment in Tucson, Arizona. Eligible participants were invited to the Collaboratory, where complete details of the study were given in the participant's preferred language (English or Spanish), and informed consent was obtained. This study was approved by the University of Arizona Institutional Review Board (approval 604536275).

### Description of GCSWLI

Participants met with a trained bilingual, bicultural Hispanic male lifestyle coach in person (one to one) for 30 to 45 minutes once a week for 12 weeks. During these individual counseling sessions, lifestyle coaches assisted participants in setting a daily calorie goal and physical activity goals (progressing to  $\geq 225$  minutes of moderate-intensity physical activity per week) and reminded participants to self-monitor in written diaries. The GCSWLI was developed based on the Diabetes Prevention Program lifestyle intervention [23]. Written materials were grounded in social cognitive theory [24] and problem-solving theory [25] and included behavior change techniques shown to be successful in weight loss interventions (eg, self-monitoring, goal setting, and self-efficacy) [26]. Materials were adapted for low literacy (fifth grade reading level) and tailored for gender and cultural factors based on formative work [15–17] and a framework developed by Bernal and Sáez-Santiago [27]. Specifically, we adapted the intervention for gender by including information related to gender role strains (the role of the man in the household) in written materials and discussed how role strains influence health behaviors [20]. This is relevant given that Hispanic culture tends to have traditional and strictly defined gender roles, which may cause differential intervention

effects for men and women [28]. To support progress toward goals, participants were required to bring their completed written diaries to counseling sessions. Challenges to meeting diet and physical activity goals were discussed, and lifestyle coaches worked with participants to identify and overcome perceived barriers (eg, strenuous work schedules) to engage in healthy behaviors.

### mHealth Intervention Components

At the conclusion of the 12-week GCSWLI study, specific mHealth components were added to the in-person GCSWLI curriculum used with WLC control participants. Specifically, the use of Fitbit Charge 2, a consumer-wearable physical activity tracker, and a Fitbit Aria Wi-Fi Smart Scale for body weight were included to complement self-monitoring using written diaries. In the first in-person session, all participants were instructed on how to use the Fitbit tools to reduce any barriers related to the introduction of mHealth technology. This included logging into their personal Fitbit dashboard, where the lifestyle coach stored and monitored the data weekly. It was recommended that participants wear the Fitbit Charge 2 during all waking hours and weigh themselves daily using the Aria smart scale. The Fitbit data were downloaded during weekly in-person sessions and discussed with the participants by the lifestyle coach. A total of 2 weekly SMS text messages were sent to participants throughout the following week by a lifestyle coach, tailored to the participants' needs based on observations from the Fitbit tools and self-reported barriers during the weekly in-person sessions. All intervention strategies from the original GCSWLI curriculum were maintained to ensure that participants had the opportunity to lose weight, independent of mHealth use.

## Quantitative Outcome Measures

Clinical assessments included anthropometrics, cardiometabolic measures, and self-reported diet and physical activity behaviors at baseline and 12 weeks. Acculturation was measured at baseline, and treatment satisfaction was measured at the completion of the study.

### Anthropometrics and Cardiometabolic Measures

Height (cm) was measured using a wall-mounted stadiometer, and weight (kg) was measured using a calibrated digital scale (Seca 876). Waist circumference (cm) was measured using a Gulick tape measure, and body composition was measured using whole-body dual-energy x-ray absorptiometry (Lunar Prodigy; Lunar). A trained phlebotomist collected fasting blood samples (venipuncture, 25 mL) using an approved protocol for examining the following cardiometabolic measures: (1) a metabolic liver panel (alanine transaminase and aspartate transaminase), (2) a lipid panel (total cholesterol, high-density lipoprotein, low-density lipoprotein, and triglycerides), (3) high-sensitivity C-reactive protein, and (4) fasting glucose and hemoglobin A<sub>1c</sub>.

### Self-reported Dietary Intake and Physical Activity

Dietary intake was assessed using the Southwestern Food Frequency Questionnaire (SWFFQ) [29]. The SWFFQ is a bilingual food frequency questionnaire adapted from the Arizona Food Frequency Questionnaire. It includes more than 150 food items, such as corn and flour tortillas, nopalitos, machaca, and chorizo, which are often consumed by individuals in southwest United States. The SWFFQ uses relevant Mexican names for food items (eg, *naranja*, not *china*, for *orange*) and asks participants to describe their average use (ranging from 3 or more times daily to rarely or never) and portion size (small, medium, or large) for each food item. Data retrieved from the SWFFQ allowed for the calculation of total daily energy intake and percentage of daily energy intake [30]. Physical activity was assessed using the validated Global Physical Activity Questionnaire [31], which asked participants to describe their activity at work, travel to and from places, recreational activities, and sedentary behavior. The questionnaire is available in both English and Spanish and provides the duration (minutes per week) and intensity of physical activity (moderate to vigorous).

### Acculturation Levels and Treatment Satisfaction

Acculturation affects attitudes, perceptions, and behaviors surrounding health-related activities and outcomes [32]. Therefore, we assessed the level of acculturation related to language, ethnic identity, and ethnic interaction in the study participants using the validated Acculturation Rating Scale for Mexican Americans-II [33], which is available in both English and Spanish. Treatment satisfaction was assessed at study completion (12 weeks) using 4 questions rated on a Likert scale, with higher scores indicating greater program favorability [34]. Questions prompted participants to rate their overall satisfaction with their progress toward changing dietary and physical activity behaviors for weight management and asked participants whether they would recommend the program to others [35].

## Qualitative Measures: Semistructured Interviews

All participants were invited to participate in individual postintervention interviews. The interview protocol was devised to elicit information about (1) participant satisfaction with the intervention, (2) adequacy of cultural and gender intervention adaptation, (3) strategies for improved intervention engagement, (4) adequacy of physical activity and dietary recommendations, and importantly, (5) perspectives on the integration of mHealth technology as a component in weight management interventions. Written and verbal consent was obtained from each willing participant, and an additional US \$25 was offered as compensation. All interviews were conducted by each participant's respective lifestyle coach in their preferred language and were audio recorded to facilitate data interpretation. To minimize bias, all interviewers carefully explained that the participants were free and under no pressure to express their views and that the focus of the interviews was to learn their perceptions to help inform the future development of similar intervention programs.

### Quantitative Statistical Analyses

Descriptive statistics were calculated for all the variables at baseline. To explore differences in baseline variables between the enrolled group and the group that did not enroll in the mHealth pilot study, we conducted 2-tailed *t* tests and chi-square tests on continuous and categorical demographic and anthropometric variables, respectively. Study attrition and process measures, including attendance at individual sessions, self-monitoring adherence, use of mHealth technology support, and treatment satisfaction, were examined using descriptive analyses. Attrition was calculated for those who did not complete 12 weeks of treatment. One-sample 2-tailed *t* tests were used to estimate the mean changes in body weight and anthropometric measures, as well as the secondary outcomes of eating behaviors, physical activity levels, and cardiometabolic biomarkers. Participants with missing blood or weight data at week 12 were excluded from the analysis. Statistical analyses were performed using SAS (version 9.4; SAS Institute Inc). *P* values of <.05 were considered statistically significant.

### Qualitative Data Analyses

All individual interviews were digitally recorded, transcribed verbatim in their respective languages (Spanish or English), and coded using NVivo (version 10, QSR International) qualitative analysis software [36]. For the purposes of this study, we analyzed only qualitative data evaluating the inclusion of mHealth technology in the weight loss intervention. As such, all interviews that elicited statements including mentions of the use of mHealth technology in the intervention were compartmentalized and included in the thematic analysis. The purpose of this analysis was to characterize participants' opinions and perspectives regarding the use of mHealth technology as a component in weight management interventions for Hispanic men.

## Results

### Overview

On completion of the initial 12-week GCSWLI, WLC participants ( $n=25$ ) were offered the mHealth-supported intervention. Overall, 2 patients were lost to follow-up for unknown reasons, 4 declined participation, and 1 individual had a measured BMI  $25 \text{ kg/m}^2$  before the start of the treatment and was deemed ineligible (Figure 1). A total of 18 participants were younger (38 vs 49 years;  $P=.04$ ) and weighed more (105 vs 90 kg;  $P=.02$ ) than those who elected not to participate. Other demographic characteristics and acculturation levels did not differ between groups (Table 1).

Overall, an average of 75% (SD 27%) of the individual sessions were attended throughout the 12-week treatment period. Of the 18 participants, 2 (11%) attended <50% of the sessions, 6 (33%) attended 50% to 75% of the sessions, and 10 (56%) attended >75% of the sessions. Of the 18 participants, 16 (89%) completed the intervention, with an overall attrition rate of 11.1%. Self-monitoring adherence and mHealth technology use at the follow-up are shown in Table 2.

On average, each person completed 6.2 weekly written diaries; 51.93% (698/1344) of the total diaries were completed. Diet was recorded on average 1.7 days per week, with an average of 1260 kcal reported each day. Participants reported physical

activity an average of 1.4 days per week, with a total of 78 minutes of weekly exercise, and they recorded their weight on an average of 1.8 days per week. Activity monitors were used 5.1 days per week, and weight was reported using wireless scales 3.2 days per week. Overall, participants wore the Fitbit 71.6% of the intervention days and weighed themselves using the smart scale for 30.5% of the days they spent in the intervention. The participants were highly satisfied with the weight management program, and 100% (15/15) of the participants indicated that they were likely to recommend the program to others. Satisfaction with overall progress toward weight loss was positive, with progress in changing dietary habits ranking the highest (Table 3).

The pre- and postweight and cardiometabolic measures are shown in Table 4. Weights were significantly reduced by an average of  $-4.7 \text{ kg}$  (95% CI  $-7.1$  to  $-2.4 \text{ kg}$ ;  $P<.001$ ); 7 individuals (44%) achieved clinically significant weight loss of  $\geq 5\%$ . In addition, significant changes in BMI by  $-1.6 \text{ kg/m}^2$  ( $-2.4$  to  $-0.8 \text{ kg/m}^2$ ), body fat by  $-1.97\%$  ( $-2.98\%$  to  $-0.96\%$ ), and waist circumference by  $-5.1 \text{ cm}$  ( $-6.8$  to  $-3.4 \text{ cm}$ ) were observed after the 12-week intervention ( $P<.001$ ). Physical activity increased, and the average caloric intake decreased, but the changes were not statistically significant. Significant decreases in cardiometabolic measures were found in hemoglobin  $A_{1c}$  ( $P=.02$ ), alanine transaminase ( $P=.04$ ), and total cholesterol ( $P=.007$ ).

**Table 1.** Demographic and baseline characteristics of participants in a mobile health (mHealth)–supported gender- and culturally sensitive weight loss intervention.

	mHealth participants (n=18)	Did not enroll (n=7)	P value
Age (years), mean (SD)	38.0 (10.9)	48.9 (12.1)	.04
BMI (kg/m <sup>2</sup> ), mean (SD)	34.3 (5.5)	30.3 (3.6) <sup>a</sup>	.14
Weight (kg), mean (SD)	104.6 (20.8)	90.0 (6.3) <sup>a</sup>	.02
Waist circumference (cm), mean (SD)	114.1 (14.6)	107.5 (5.7) <sup>a</sup>	.34
Employed, n (%)	12 (66)	6 (85)	.63
<b>Income (US \$), n (%)</b>			.99
<30,000	13 (72)	5 (71)	
30,000-60,000	4 (22)	2 (29)	
>60,000	1 (6)	0 (0)	
<b>Education, n (%)</b>			.99
Less than high school	6 (33)	2 (29)	
High school or General Educational Development	5 (28)	2 (29)	
Greater than high school	7 (39)	3 (43)	
Married or lives with a domestic partner, n (%)	14 (78)	7 (100)	.30
US born, n (%)	10 (56)	1 (14)	.09
<b>Heritage, n (%)</b>			.53
Mexican	10 (56)	6 (86)	
Mexican American	7 (39)	1 (14)	
Puerto Rican	1 (6)	0 (0)	
<b>Primary language spoken at home, n (%)</b>			.39
Spanish	10 (56)	6 (86)	
English	6 (33)	1 (14)	
Both equally	2 (11)	0 (0)	
Mexican Orientation Subscale of ARSMA-II <sup>b</sup> , mean (SD)	65.6 (16.7)	73.6 (8.2)	.24
<b>Acculturation level (ARSMA-II), n (%)</b>			.79
Very Mexican oriented	7 (39)	5 (71)	
Mexican oriented to approximately balanced to oriented bicultural	6 (33)	1 (14)	
Slightly Anglo bicultural	3 (17)	1 (14)	
Strongly Anglo oriented	1 (6)	0 (0)	
Very assimilated	1 (6)	0 (0)	

<sup>a</sup>A total of 2 individuals dropped out owing to their participation in the waitlist control.

<sup>b</sup>ARSMA-II: Acculturation Rating Scale for Mexican Americans-II.

**Table 2.** Attendance, self-monitoring adherence, and mobile health (mHealth) technology use in a gender- and culturally sensitive weight loss intervention (N=18).

Characteristics	Values
<b>Overall percentage of attendance at individual sessions, mean (SD)</b>	75.0 (27.0)
Attended <50% of the sessions, n (%)	2 (11)
Attended 50%-75% of the sessions, n (%)	6 (33)
Attended >75% of the sessions, n (%)	10 (56)
<b>Self-reported measures<sup>a</sup></b>	
Diaries completed, n (%)	112 (52)
Diaries completed per person, mean (SD)	6.2 (4.2)
Diet days recorded (days per week), mean (SD)	1.7 (2.8)
Self-reported calorie intake (kcal per day), mean (SD)	1260 (566)
<b>Self-reported physical activity , mean (SD)</b>	
Days per week	1.4 (2.0)
Minutes per week	78.0 (63.1)
Self-weighed (days per week), mean (SD)	1.8 (2.9)
<b>mHealth measures, mean (SD)</b>	
Activity monitor use (days per week)	5.1 (2.0)
Activity monitor use during sleep (days per week)	2.2 (2.4)
Diet days recorded via web (days per week)	0.4 (1.3)
Weight reported using wireless scale (days per week)	3.2 (3.2)

<sup>a</sup>Data obtained from paper diary logging.

**Table 3.** Participant responses to treatment satisfaction survey (for 15 of 16 completers).

Participant responses	Values
<b>How satisfied are you overall with the weight management program?<sup>a</sup> n (%)</b>	
Somewhat satisfied	1 (7)
Very satisfied	13 (87)
Missing response	1 (7)
Would you recommend the weight management program you received to others? <sup>b</sup> (definitely would), n (%)	15 (100)
Given the effort you put into following the weight management program, how satisfied are you with your overall progress over the past 12 weeks? <sup>c</sup> median (IQR)	3.0 (2.5-4.0)
<b>Given the effort you put into following the weight management program for the past 12 weeks, how satisfied are you overall with your progress on?<sup>c</sup> median (IQR)</b>	
Changing your weight	3.0 (3.0-4.0)
Changing your dietary habits	3.0 (2.5-4.0)
Changing your physical activity habits	3.0 (2.0-4.0)

<sup>a</sup>1=very dissatisfied and 4=very satisfied.

<sup>b</sup>1=definitely not and 4=definitely would.

<sup>c</sup>4=very dissatisfied and 4=very satisfied.

**Table 4.** Outcomes before and after the mobile health intervention and estimated mean change from baseline.

	Baseline (n=18), mean (SD)	Week 12 (n=16), mean (SD)	Change from baseline, mean (95% CI)	P value
Weight (kg)	104.6 (20.8)	100.3 (21.6)	-4.7 (-7.1 to -2.4)	<.001
BMI (kg/m <sup>2</sup> )	34.3 (5.5)	32.7 (5.9)	-1.6 (-2.4 to -0.8)	<.001
Percentage weight change	— <sup>a</sup>	-4.8 (4.5)	-4.8 (-7.0 to -2.0)	<.001
Percentage body fat	37.3 (3.6)	35.3 (4.8)	-1.97 (-2.98 to -0.96)	<.001
Waist circumference (cm)	114.1 (14.6)	109.5 (14.3)	-5.1 (-6.8 to -3.4)	<.001
Leisure-time physical activity (min/week)	87.2 (100.6)	139.4 (170.8)	60.0 (-44.6 to 164.6)	.24
Average caloric intake (kcal/day)	3102 (3745)	1754 (1329)	-1196 (-2647 to 254)	.11
<b>Cardiometabolic measures</b>				
Hemoglobin A <sub>1c</sub> (%)	5.6 (0.8)	5.5 (0.6)	-0.22 (-0.41 to -0.03)	.02
Glucose (mg/100 mL)	105.3 (28.4)	105.4 (32.8)	-2.31 (-9.26 to 4.64)	.49
Alanine transaminase (U/L)	53.6 (52.7)	39.6 (28.1)	-16.9 (-32.5 to -1.23)	.04
Aspartate transaminase (U/L)	28.9 (19.3)	24.9 (9.6)	-5.3 (-11.7 to 1.04)	.10
Total cholesterol (mg/100 mL)	185.7 (40.4)	171.4 (37.5)	-11.8 (-19.9 to -3.7)	.007
High-density lipoprotein cholesterol (mg/100 mL)	40.2 (5.6)	39.3 (6.9)	-1.0 (-4.2 to 2.2)	.51
Low-density lipoprotein cholesterol (mg/100 mL)	113.7 (34.3)	104.8 (31.8)	-5.6 (-12.0 to 0.80)	.08
Triglycerides (mg/100 mL)	158.8 (82.2)	136.4 (71.5)	-25.6 (-52.0 to 0.9)	.06
High-sensitivity C-reactive protein (mg/100 mL)	3.4 (2.9)	2.5 (2.5)	-0.83 (-2.37 to 0.70)	.26

<sup>a</sup>Not available.

### Participant Experience With mHealth Components

Of the 16 participants who completed the intervention, 14 (88%) agreed to participate in an individual interview after the completion of the 12-week intervention. A total of 4 overarching themes emerged during these interviews regarding participants' perceptions of the integration of mHealth technology into a weight management intervention.

When asked to reflect on their use of Fitbit activity and weight trackers, participants expressed positive feedback and no concerns about the security or privacy of the e-supported data were reported. The participants explained that mHealth technology was at times easier to use than the written diary self-monitoring method. This, in turn, facilitated the seamless integration of self-monitoring into daily lifestyle behavior changes. For example, a participant commented the following:

*And of course the best part that I really enjoyed were the technological tools, the Fitbit the scale things like that the Fitbit app that made a huge world of difference. You know it became part of my life and now I use it all the time to count my steps look at information look at data look at how I sleep and how much exercise I got that day it's great it's a great thing to have.*

Participants also expressed that using the Fitbit wearable tracking device facilitated accountability to daily goals.

Specifically, the participants expressed that it was easy to see when daily goals were not being met and that they enjoyed the positive feedback they received when goals were met. A participant stated the following:

*[I liked the Fitbit] because I can have a record of whether or not I walk, you understand me? Instead of forgetting...and when I do wrap up a night, it will tell me, "hey great job! You walked like 10, 20, 25 minutes" and I pay attention to that. So that helps.*

Another participant commented on the utility of activity trackers to standardize step count goals, explaining as follows:

*I like [the Fitbit] because, how do I explain? I can at least say that 10,000 steps is like the bottom line, right?...So, when I realize that I have not reached 3,000 steps for two or three days I really start paying attention.*

Conversely, some participants expressed concerns about the use of technology. For some, reliable internet connectivity was a barrier to the use of supplemental technology. The Fitbit scale requires a reliable internet connection to sync and upload data, which some participants did not have access to. A few men also explained that an additional technology-centered component was an added burden instead of a supplemental benefit of the intervention. This was particularly the case with older participants, who reported less technology literacy because

technology was not already integrated into their daily lives. A participant described the following:

*I think that was helpful. I don't know that the Fitbit thing would have been very helpful to me I'm not a real tech-oriented guy I don't want to sync up this and that enter info, yeah so I'm not, I'm not into that so I don't know if that would have helped me, but other things did that's for sure.*

## Discussion

### Principal Findings

Hispanic men are part of the largest racial and ethnic subpopulation in the United States and are disproportionately affected by obesity-related chronic diseases [37]. Although broad systemic sociopolitical and sociocultural factors have an undeniable and ubiquitous influence on this burden, disparities are also partly owing to the lack of tailored culturally and linguistically responsive programs. The overarching objective of this pilot study was to examine the feasibility and acceptability of integrating mHealth technology into a 12-week GCSWLI for Hispanic men with overweight and obesity. The findings of this work help elucidate the role that mHealth technologies can play in supporting the delivery of preventive and interventional care to at-risk and hard-to-reach groups, including Hispanic men of Mexican origin.

### Comparison With Prior Work

The findings of this study yielded promising improvements in body composition, including changes in body weight, body fat percentage, and waist circumference. Recent literature assessing body composition outcomes has yielded only modest results in other Hispanic subpopulations and has been largely limited to enrollment of women [38,39]. In fact, only a few studies have focused solely on Hispanic men. In the previously mentioned ANIMO study, where mHealth technology was not included, the mean weight loss in the GCSWLI at week 12 was  $-6.3$  kg (95% CI  $-8.1$  to  $-4.4$  kg), and it was maintained at week 24 ( $-6.4$  kg, 95% CI  $-8.6$  to  $-4.3$  kg) [21]. More recently, Rosas et al [14] conducted a study focusing on a choice-based technology-mediated lifestyle intervention for 200 Latino men. It compared 3 weight loss intervention options (videoconference, web-based videos, and in person) based on participants' choices. Overall, 41% (82/200) chose the in-person group, 31% (62/200) chose the web-based video group, and 28% (56/200) chose the videoconference group. The authors found that participants who initially chose the videoconference and in-person group sessions lost significantly more weight at 6 months (mean  $-3.9$ , SD 6.1 kg for videoconference and mean  $-4.3$ , SD 5.3 kg for in-person sessions) compared with web-based videos (mean  $-0.3$ , SD 3.7 kg). Similar weight loss was observed for our mHealth-supported GCSWLI at week 12 ( $-4.7$  kg, 95% CI  $-7.1$  to  $-2.4$  kg). Notably, attendance was high across all 3 studies, independent of the intervention delivery method. For example, attendance at individual counseling sessions for ANIMO was 77% and was 75% (144/192) for this mHealth study. Rosas et al [14] observed that the mean number of sessions attended out of 12 weeks was 10.9 (SD 2.6) for the videoconference sessions and 9.7 (SD 4.3) for in-person sessions. Continued efforts to

examine which aspects of these studies, including delivery options and which intervention components (eg, mHealth) influence study effectiveness, are warranted.

Regarding feasibility and acceptability, our mHealth intervention was well received and produced a retention rate of 88.9%. In general, weight loss interventions in Hispanic adults have previously reported lower retention rates; however, they did not include mHealth components [40]. Rosas et al [14] observed a retention rate of 96.9%, suggesting that technology-mediated strategies may support retention efforts [14]. Our findings suggest moderate adherence to mHealth technology prescriptions, as the Fitbit Charge 2 wearable activity monitor was used 71.58% (962/1344) of the intervention days, and the Fitbit Aria Wi-Fi scale was used 30.51% (410/1344) of the intervention days. However, the completion rate of diaries (698/1344, 51.93%) was lower than that in the ANIMO study (70%) [21]. This suggests that adding mHealth technology decreased adherence to self-monitoring behaviors. For example, diet behaviors were reported only 1.7 days per week, with an average of 1260 kcal reported each day, which was likely underreported. However, our findings align with a recent systematic review that suggests wearable tracking devices as part of weight loss interventions appear to be feasible when incorporated in short-term (<6 months) comprehensive weight loss programs [41]. We found that using mHealth tools was well received in a 12-week weight loss intervention for Hispanic men. Specifically, the qualitative findings suggested that participants' perceptions on the use of mHealth technology specific to this study were acceptable and facilitated accountability to lifestyle behaviors during the study.

Hispanic men experienced high levels of overall satisfaction with this intervention modality expressing that they would recommend participation to other Hispanic men. This is in agreement with the findings of Wang et al [9] who confirmed that wearable technologies and apps appear to have high efficacy when used in the context of medical care and coaching for weight loss. Nevertheless, a few participants cited internet connection as a barrier to the use of the Aria Scale for weighing themselves daily. This may have affected our findings, as daily weighing has been shown to improve weight loss and adoption of weight control behaviors [42]. This finding aligns with other studies that have cited internet accessibility as a barrier to intervention delivery [8]. Mayberry et al [8] reported in their findings that internet access remains unavailable to multiple segments of disadvantaged or vulnerable populations, including people with low socioeconomic status, members of racial and ethnic groups, and persons with limited health literacy or numeracy skills. Future research focusing on the incorporation of technology tracking devices into comprehensive weight loss interventions should prioritize the provision of reliable internet connectivity to participants. This can ensure that these devices fulfill their purpose of complementing participants' efforts for behavior change instead of becoming an obstacle to their goals.

### Limitations

Although this study has many strengths, we must acknowledge some limitations, including a small study sample, a short time frame, and a 1-group pre-post design that did not allow direct

comparison with a control group. Therefore, we cannot disentangle which mHealth components were most supportive of weight loss, and the results must be interpreted with caution. Furthermore, although most individuals achieved weight loss, we did not observe statistically significant changes in diet or physical activity. Recall bias should be considered as a hindering factor when self-reporting techniques are involved. However, the investigative team attempted to minimize recall bias by reviewing the questionnaires in person for each participant. Furthermore, given that all study participants came from Tucson, Arizona, and were largely of Mexican-origin descent apart from one individual, this decreases the generalizability of findings and adaptability to Hispanic men in other areas. There is also a possibility that increased adherence to mHealth technology could have been due to the novelty of using the Fitbit wearable technology and gaining access to high-level health care tools such as dual-energy x-ray absorptiometry scan and cardiometabolic measures. Future research in this area should explore the impact of this intervention modality beyond 12 weeks to explore the progression of healthy behavior continuity and adherence levels to wearable technology in the long term.

In addition, interventions should focus on underrepresented and underresourced subpopulations that are disproportionately affected by health inequalities and have unequal access to high-quality health services and information.

### Conclusions

The integration of mHealth technology into our 12-week GCSWLI appeared to be feasible and widely accepted by study participants. Although the use of wearable technology was modest, Hispanic men achieved clinically meaningful weight loss at the end of the intervention. This pilot study contributes to the growing literature on health promotion and mHealth technologies to aid the adoption of healthy lifestyle behavior changes implemented through weight loss interventions for Hispanic men. Research efforts centered on weight management, and mHealth should consider culturally adapted frameworks that address barriers hindering the awareness and maintenance of healthy lifestyles. The use of mHealth holds promise in lifestyle interventions for Hispanic men, making healthy living opportunities more practical and achievable for communities that have historically lacked the benefits of this type of technology.

### Acknowledgments

This work was supported by the University of Arizona Cancer Center Disparities Pilot Project Award, the University of Arizona Foundation, and Dean's Canyon Ranch Center for Prevention and Health Promotion Fund.

### Authors' Contributions

DOG, LAV, MLB, and SPH designed the research. DOG, LAV, and BA conducted the research. MLB, BAR, LAV, and DOG analyzed the data. All authors were involved in writing the paper and provided final approval of the submitted and published versions.

### Conflicts of Interest

None declared.

### References

1. Colby SL, Ortman JM. Projections of the Size and Composition of the U.S. Population: 2014 to 2060. United States Census Bureau. 2015 Mar. URL: <https://www.census.gov/content/dam/Census/library/publications/2015/demo/p25-1143.pdf> [accessed 2022-09-12]
2. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017-2018. NCHS Data Brief 2020 Feb(360):1-8 [FREE Full text] [Medline: [32487284](#)]
3. Cowie CC, Casagrande SS, Geiss LS. Prevalence and incidence of type 2 diabetes and prediabetes. In: Cowie CC, Casagrande SS, Menke A, Cissell MA, Eberhardt MS, Meigs JB, et al, editors. Diabetes in America. 3rd edition. Bethesda, MD, USA: National Institute of Diabetes and Digestive and Kidney Diseases (US); 2018.
4. Kallwitz ER, Daviglius ML, Allison MA, Emory KT, Zhao L, Kuniholm MH, et al. Prevalence of suspected nonalcoholic fatty liver disease in Hispanic/Latino individuals differs by heritage. Clin Gastroenterol Hepatol 2015 Mar;13(3):569-576 [FREE Full text] [doi: [10.1016/j.cgh.2014.08.037](#)] [Medline: [25218670](#)]
5. Franz MJ, Boucher JL, Rutten-Ramos S, VanWormer JJ. Lifestyle weight-loss intervention outcomes in overweight and obese adults with type 2 diabetes: a systematic review and meta-analysis of randomized clinical trials. J Acad Nutr Diet 2015 Sep;115(9):1447-1463. [doi: [10.1016/j.jand.2015.02.031](#)] [Medline: [25935570](#)]
6. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: [10.1007/s12160-016-9846-0](#)] [Medline: [27757789](#)]
7. 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](#)]



8. Mayberry LS, Lyles CR, Oldenburg B, Osborn CY, Parks M, Peek ME. mHealth interventions for disadvantaged and vulnerable people with type 2 diabetes. *Curr Diab Rep* 2019 Nov 25;19(12):148 [FREE Full text] [doi: [10.1007/s11892-019-1280-9](https://doi.org/10.1007/s11892-019-1280-9)] [Medline: [31768662](https://pubmed.ncbi.nlm.nih.gov/31768662/)]
9. Wang E, Abrahamson K, Liu PJ, Ahmed A. Can mobile technology improve weight loss in overweight adults? A systematic review. *West J Nurs Res* 2020 Sep;42(9):747-759. [doi: [10.1177/0193945919888224](https://doi.org/10.1177/0193945919888224)] [Medline: [31762402](https://pubmed.ncbi.nlm.nih.gov/31762402/)]
10. Arora S, Ford K, Terp S, Abramson T, Ruiz R, Camilon M, et al. Describing the evolution of mobile technology usage for Latino patients and comparing findings to national mHealth estimates. *J Am Med Inform Assoc* 2016 Sep;23(5):979-983 [FREE Full text] [doi: [10.1093/jamia/ocv203](https://doi.org/10.1093/jamia/ocv203)] [Medline: [26995564](https://pubmed.ncbi.nlm.nih.gov/26995564/)]
11. Godino JG, Golaszewski NM, Norman GJ, Rock CL, Griswold WG, Arredondo E, et al. Text messaging and brief phone calls for weight loss in overweight and obese English- and Spanish-speaking adults: a 1-year, parallel-group, randomized controlled trial. *PLoS Med* 2019 Sep 25;16(9):e1002917 [FREE Full text] [doi: [10.1371/journal.pmed.1002917](https://doi.org/10.1371/journal.pmed.1002917)] [Medline: [31553725](https://pubmed.ncbi.nlm.nih.gov/31553725/)]
12. Rosas LG, Lv N, Xiao L, Lewis MA, Zavella P, Kramer MK, et al. Evaluation of a culturally-adapted lifestyle intervention to treat elevated cardiometabolic risk of Latino adults in primary care (Vida Sana): a randomized controlled trial. *Contemp Clin Trials* 2016 May;48:30-40 [FREE Full text] [doi: [10.1016/j.cct.2016.03.003](https://doi.org/10.1016/j.cct.2016.03.003)] [Medline: [26995280](https://pubmed.ncbi.nlm.nih.gov/26995280/)]
13. Rosas LG, Lv N, Xiao L, Lewis MA, Venditti EM, Zavella P, et al. Effect of a culturally adapted behavioral intervention for Latino adults on weight loss over 2 years: a randomized clinical trial. *JAMA Netw Open* 2020 Dec 01;3(12):e2027744 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.27744](https://doi.org/10.1001/jamanetworkopen.2020.27744)] [Medline: [33337491](https://pubmed.ncbi.nlm.nih.gov/33337491/)]
14. Rosas LG, Lv N, Xiao L, Azar KM, Hooker SP, Venditti EM, et al. Preferences for technology-mediated behavioral lifestyle interventions with different levels of coach and peer support among Latino men: comparative study within one arm of a randomized controlled trial. *JMIR Form Res* 2022 Feb 04;6(2):e29537 [FREE Full text] [doi: [10.2196/29537](https://doi.org/10.2196/29537)] [Medline: [35119377](https://pubmed.ncbi.nlm.nih.gov/35119377/)]
15. Garcia DO, Valdez LA, Hooker SP. Hispanic male's perspectives of health behaviors related to weight management. *Am J Mens Health* 2017 Sep;11(5):1547-1559 [FREE Full text] [doi: [10.1177/1557988315619470](https://doi.org/10.1177/1557988315619470)] [Medline: [26634854](https://pubmed.ncbi.nlm.nih.gov/26634854/)]
16. Valdez LA, Amezquita A, Hooker SP, Garcia DO. Mexican-origin male perspectives of diet-related behaviors associated with weight management. *Int J Obes (Lond)* 2017 Dec;41(12):1824-1830. [doi: [10.1038/ijo.2017.173](https://doi.org/10.1038/ijo.2017.173)] [Medline: [28757642](https://pubmed.ncbi.nlm.nih.gov/28757642/)]
17. Valdez LA, Morrill KE, Griffith DM, Lindberg NM, Hooker SP, Garcia DO. Mexican origin Hispanic men's perspectives of physical activity-related health behaviors. *Am J Mens Health* 2019;13(1):1557988319834112 [FREE Full text] [doi: [10.1177/1557988319834112](https://doi.org/10.1177/1557988319834112)] [Medline: [30819068](https://pubmed.ncbi.nlm.nih.gov/30819068/)]
18. Valdez LA, Garcia DO. Hispanic male recruitment into obesity-related research: evaluating content messaging strategies, experimental findings, and practical implications. *Int Q Community Health Educ* 2021 Oct;42(1):85-93. [doi: [10.1177/0272684X20982598](https://doi.org/10.1177/0272684X20982598)] [Medline: [33356864](https://pubmed.ncbi.nlm.nih.gov/33356864/)]
19. Pagoto SL, Schneider KL, Oleski JL, Luciani JM, Bodenlos JS, Whited MC. Male inclusion in randomized controlled trials of lifestyle weight loss interventions. *Obesity (Silver Spring)* 2012 Jun;20(6):1234-1239 [FREE Full text] [doi: [10.1038/oby.2011.140](https://doi.org/10.1038/oby.2011.140)] [Medline: [21633403](https://pubmed.ncbi.nlm.nih.gov/21633403/)]
20. Garcia DO, Valdez LA, Bell ML, Humphrey K, Hingle M, McEwen M, et al. A gender- and culturally-sensitive weight loss intervention for Hispanic males: the ANIMO randomized controlled trial pilot study protocol and recruitment methods. *Contemp Clin Trials Commun* 2018 Mar;9:151-163 [FREE Full text] [doi: [10.1016/j.conctc.2018.01.010](https://doi.org/10.1016/j.conctc.2018.01.010)] [Medline: [29696238](https://pubmed.ncbi.nlm.nih.gov/29696238/)]
21. Garcia DO, Valdez LA, Aceves B, Bell ML, Humphrey K, Hingle M, et al. A gender- and culturally sensitive weight loss intervention for Hispanic men: results from the animo pilot randomized controlled trial. *Health Educ Behav* 2019 Oct;46(5):763-772. [doi: [10.1177/1090198119853011](https://doi.org/10.1177/1090198119853011)] [Medline: [31216875](https://pubmed.ncbi.nlm.nih.gov/31216875/)]
22. Weir CB, Jan A. BMI Classification Percentile And Cut Off Points. Treasure Island, FL, USA: StatPearls Publishing; 2022.
23. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. *Diabetes Care* 2002 Dec;25(12):2165-2171 [FREE Full text] [doi: [10.2337/diacare.25.12.2165](https://doi.org/10.2337/diacare.25.12.2165)] [Medline: [12453955](https://pubmed.ncbi.nlm.nih.gov/12453955/)]
24. Bandura A. Social cognitive theory: an agentic perspective. *Annu Rev Psychol* 2001;52:1-26. [doi: [10.1146/annurev.psych.52.1.1](https://doi.org/10.1146/annurev.psych.52.1.1)] [Medline: [11148297](https://pubmed.ncbi.nlm.nih.gov/11148297/)]
25. D'Zurilla TJ, Goldfried MR. Problem solving and behavior modification. *J Abnorm Psychol* 1971 Aug;78(1):107-126. [doi: [10.1037/h0031360](https://doi.org/10.1037/h0031360)] [Medline: [4938262](https://pubmed.ncbi.nlm.nih.gov/4938262/)]
26. Michie S, West R, Sheals K, Godinho CA. Evaluating the effectiveness of behavior change techniques in health-related behavior: a scoping review of methods used. *Transl Behav Med* 2018 Mar 01;8(2):212-224 [FREE Full text] [doi: [10.1093/tbm/ibx019](https://doi.org/10.1093/tbm/ibx019)] [Medline: [29381786](https://pubmed.ncbi.nlm.nih.gov/29381786/)]
27. Bernal G, Sáez-Santiago E. Culturally centered psychosocial interventions. *J Community Psychol* 2006 Mar;34(2):121-132. [doi: [10.1002/jcop.20096](https://doi.org/10.1002/jcop.20096)]
28. Rosas LG, Thiyagarajan S, Goldstein BA, Drieling RL, Romero PP, Ma J, et al. The effectiveness of two community-based weight loss strategies among obese, low-income US Latinos. *J Acad Nutr Diet* 2015 Apr;115(4):537-50.e2 [FREE Full text] [doi: [10.1016/j.jand.2014.10.020](https://doi.org/10.1016/j.jand.2014.10.020)] [Medline: [25578925](https://pubmed.ncbi.nlm.nih.gov/25578925/)]

29. Garcia RA, Taren D, Teufel NI. Factors associated with the reproducibility of specific food items from the Southwest Food Frequency Questionnaire. *Ecol Food Nutrition* 2000 Jan;38(6):549-561. [doi: [10.1080/03670244.2000.9991596](https://doi.org/10.1080/03670244.2000.9991596)]
30. Taren D, de Tobar M, Ritenbaugh C, Graver E, Whitacre R, Aickin M. Evaluation of the southwest food frequency questionnaire. *Ecol Food Nutrition* 2000 Jan;38(6):515-547. [doi: [10.1080/03670244.2000.9991595](https://doi.org/10.1080/03670244.2000.9991595)]
31. Cleland CL, Hunter RF, Kee F, Cupples ME, Sallis JF, Tully MA. Validity of the global physical activity questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. *BMC Public Health* 2014 Dec 10;14:1255 [FREE Full text] [doi: [10.1186/1471-2458-14-1255](https://doi.org/10.1186/1471-2458-14-1255)] [Medline: [25492375](https://pubmed.ncbi.nlm.nih.gov/25492375/)]
32. Allen JD, Caspi C, Yang M, Leyva B, Stoddard AM, Tamers S, et al. Pathways between acculturation and health behaviors among residents of low-income housing: the mediating role of social and contextual factors. *Soc Sci Med* 2014 Dec;123:26-36 [FREE Full text] [doi: [10.1016/j.socscimed.2014.10.034](https://doi.org/10.1016/j.socscimed.2014.10.034)] [Medline: [25462602](https://pubmed.ncbi.nlm.nih.gov/25462602/)]
33. Cuellar I, Arnold B, Maldonado R. Acculturation rating scale for Mexican Americans-II: a revision of the original ARSMA scale. *Hispanic J Behav Sci* 1995 Aug 1;17(3):275-304. [doi: [10.1177/07399863950173001](https://doi.org/10.1177/07399863950173001)]
34. VanWormer JJ, Martinez AM, Cosentino D, Pronk NP. Satisfaction with a weight loss program: what matters? *Am J Health Promot* 2010;24(4):238-245. [doi: [10.4278/ajhp.080613-QUAN-92](https://doi.org/10.4278/ajhp.080613-QUAN-92)] [Medline: [20232605](https://pubmed.ncbi.nlm.nih.gov/20232605/)]
35. Baldwin AS, Rothman AJ, Jeffery RW. Satisfaction with weight loss: examining the longitudinal covariation between people's weight-loss-related outcomes and experiences and their satisfaction. *Ann Behav Med* 2009 Dec;38(3):213-224 [FREE Full text] [doi: [10.1007/s12160-009-9148-x](https://doi.org/10.1007/s12160-009-9148-x)] [Medline: [20082163](https://pubmed.ncbi.nlm.nih.gov/20082163/)]
36. Edwards-Jones A. Qualitative data analysis with NVIVO. *J Educ Teach* 2014 Jan 13;40(2):193-195. [doi: [10.1080/02607476.2013.866724](https://doi.org/10.1080/02607476.2013.866724)]
37. Velasco-Mondragon E, Jimenez A, Palladino-Davis AG, Davis D, Escamilla-Cejudo JA. Hispanic health in the USA: a scoping review of the literature. *Public Health Rev* 2016 Dec 7;37:31 [FREE Full text] [doi: [10.1186/s40985-016-0043-2](https://doi.org/10.1186/s40985-016-0043-2)] [Medline: [29450072](https://pubmed.ncbi.nlm.nih.gov/29450072/)]
38. Vieira ER, Cavalcanti FA, Civitella F, Hollifield M, Caceres S, Carreno J, et al. Effects of exercise and diet on body composition and physical function in older Hispanics with type 2 diabetes. *Int J Environ Res Public Health* 2021 Jul 29;18(15):8019 [FREE Full text] [doi: [10.3390/ijerph18158019](https://doi.org/10.3390/ijerph18158019)] [Medline: [34360312](https://pubmed.ncbi.nlm.nih.gov/34360312/)]
39. Lopez DI, Chacon L, Vasquez D, Brown LD. Body composition outcomes of Healthy Fit and the role of acculturation among low-income Hispanics on the US-Mexico border. *BMC Public Health* 2021 May 25;21(1):976 [FREE Full text] [doi: [10.1186/s12889-021-11015-0](https://doi.org/10.1186/s12889-021-11015-0)] [Medline: [34034711](https://pubmed.ncbi.nlm.nih.gov/34034711/)]
40. Morrill KE, Lopez-Pentecost M, Molina L, Pfander JL, Hingle MD, Klimentidis YC, et al. Weight loss interventions for Hispanic women in the United States: a systematic review. *J Environ Public Health* 2021 Aug 19;2021:8714873 [FREE Full text] [doi: [10.1155/2021/8714873](https://doi.org/10.1155/2021/8714873)] [Medline: [34457010](https://pubmed.ncbi.nlm.nih.gov/34457010/)]
41. Cheatham SW, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. *J Sports Med Phys Fitness* 2018 Apr;58(4):534-548. [doi: [10.23736/S0022-4707.17.07437-0](https://doi.org/10.23736/S0022-4707.17.07437-0)] [Medline: [28488834](https://pubmed.ncbi.nlm.nih.gov/28488834/)]
42. Steinberg DM, Bennett GG, Askew S, Tate DF. Weighing every day matters: daily weighing improves weight loss and adoption of weight control behaviors. *J Acad Nutr Diet* 2015 Apr;115(4):511-518 [FREE Full text] [doi: [10.1016/j.jand.2014.12.011](https://doi.org/10.1016/j.jand.2014.12.011)] [Medline: [25683820](https://pubmed.ncbi.nlm.nih.gov/25683820/)]

## Abbreviations

**GCSWLI:** gender- and culturally sensitive weight loss intervention

**mHealth:** mobile health

**SWFFQ:** Southwestern Food Frequency Questionnaire

**WLC:** waitlist control

*Edited by A Mavragani; submitted 04.03.22; peer-reviewed by L Rosas, M Savolainen, H Ranjani; comments to author 24.06.22; revised version received 12.08.22; accepted 19.08.22; published 21.09.22.*

### *Please cite as:*

Garcia DO, Valdez LA, Aceves B, Bell ML, Rabe BA, Villavicencio EA, Marrero DG, Melton F, Hooker SP  
*mHealth-Supported Gender- and Culturally Sensitive Weight Loss Intervention for Hispanic Men With Overweight and Obesity: Single-Arm Pilot Study*  
*JMIR Form Res* 2022;6(9):e37637  
URL: <https://formative.jmir.org/2022/9/e37637>  
doi: [10.2196/37637](https://doi.org/10.2196/37637)  
PMID: [36129735](https://pubmed.ncbi.nlm.nih.gov/36129735/)

©David O Garcia, Luis A Valdez, Benjamin Aceves, Melanie L Bell, Brooke A Rabe, Edgar A Villavicencio, David G Marrero, Forest Melton, Steven P Hooker. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 21.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Feasibility of Measuring Screen Time, Activity, and Context Among Families With Preschoolers: Intensive Longitudinal Pilot Study

Hannah Parker<sup>1</sup>, MSc; Sarah Burkart<sup>1</sup>, PhD; Layton Reesor-Oyer<sup>1</sup>, PhD; Michal T Smith<sup>1</sup>, MSc; Roddrick Dugger<sup>1</sup>, MPH; Lauren von Klingraeff<sup>1</sup>, MPH; R Glenn Weaver<sup>1</sup>, PhD; Michael W Beets<sup>1</sup>, PhD; Bridget Armstrong<sup>1</sup>, PhD

Department of Exercise Science, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

**Corresponding Author:**

Bridget Armstrong, PhD  
Department of Exercise Science  
Arnold School of Public Health  
University of South Carolina  
921 Assembly St  
Columbia, SC, 29208-3904  
United States  
Phone: 1 803 576 8418  
Email: [ba12@mailbox.sc.edu](mailto:ba12@mailbox.sc.edu)

## Abstract

**Background:** Digital media has made screen time more available across multiple contexts, but our understanding of the ways children and families use digital media has lagged behind the rapid adoption of this technology.

**Objective:** This study evaluated the feasibility of an intensive longitudinal data collection protocol to objectively measure digital media use, physical activity, sleep, sedentary behavior, and socioemotional context among caregiver-child dyads. This paper also describes preliminary convergent validity of ecological momentary assessment (EMA) measures and preliminary agreement between caregiver self-reported phone use and phone use collected from passive mobile sensing.

**Methods:** Caregivers and their preschool-aged child (3-5 years) were recruited to complete a 30-day assessment protocol. Within 30-days, caregivers completed 7 days of EMA to measure child behavior problems and caregiver stress. Caregivers and children wore an Axivity AX3 (Newcastle Upon Tyne) accelerometer to assess physical activity, sedentary behavior, and sleep. Phone use was assessed via passive mobile sensing; we used Chronicle for Android users and screenshots of iOS screen time metrics for iOS users. Participants were invited to complete a second 14-day protocol approximately 3-12 months after their first assessment. We used Pearson correlations to examine preliminary convergent validity between validated questionnaire measures of caregiver psychological functioning, child behavior, and EMA items. Root mean square errors were computed to examine the preliminary agreement between caregiver self-reported phone use and objective phone use.

**Results:** Of 110 consenting participants, 105 completed all protocols (105/110, 95.5% retention rate). Compliance was defined a priori as completing  $\geq 70\%$ -75% of each protocol task. There were high compliance rates for passive mobile sensing for both Android (38/40, 95%) and iOS (64/65, 98%). EMA compliance was high (105/105, 100%), but fewer caregivers and children were compliant with accelerometry (62/99, 63% and 40/100, 40%, respectively). Average daily phone use was 383.4 (SD 157.0) minutes for Android users and 354.7 (SD 137.6) minutes for iOS users. There was poor agreement between objective and caregiver self-reported phone use; root mean square errors were 157.1 and 81.4 for Android and iOS users, respectively. Among families who completed the first assessment, 91 re-enrolled to complete the protocol a second time, approximately 7 months later (91/105, 86.7% retention rate).

**Conclusions:** It is feasible to collect intensive longitudinal data on objective digital media use simultaneously with accelerometry and EMA from an economically and racially diverse sample of families with preschool-aged children. The high compliance and retention of the study sample are encouraging signs that these methods of intensive longitudinal data collection can be completed in a longitudinal cohort study. The lack of agreement between self-reported and objectively measured mobile phone use highlights the need for additional research using objective methods to measure digital media use.

**International Registered Report Identifier (IRRID):** RR2-36240

(*JMIR Form Res* 2022;6(9):e40572) doi:[10.2196/40572](https://doi.org/10.2196/40572)

**KEYWORDS**

ecological momentary assessment; accelerometry; objective digital media use; screen time; sleep; activity; preschool; dyads; mobile phone

## Introduction

### Background

Excessive screen time for children is linked with poor sleep, inactivity, and behavior problems [1-4]. Few preschool-aged children meet the World Health Organization's recommendation  $\leq 1$  hour of screen time per day [5,6]. This is partly attributable to the rapid growth of digital media technology, which makes screens more available across multiple contexts [7,8]. Unfortunately, our understanding of the unique ways in which children and families use digital media has lagged behind the rapid adoption of this technology. Therefore, we need updated paradigms to understand how families use digital media and how it impacts their health.

Research on digital media use (defined here as tablet or mobile phone use) has been hindered by methodological limitations, including a reliance on retrospective self-reported measures (which are subject to recall and desirability biases) [9-11] and a tendency to overlook momentary etiological processes occurring within each day [10]. The limited research on within-day time-varying contexts (ie, child behavior and parenting stress) that influence digital media use in specific situations leads to a coarse understanding of the mechanisms underlying the link between screen time and behavioral health outcomes.

Parents allow children to use digital media for a variety of reasons (eg, to provide relief from caregiving, and to modify behavior) [12], and the decisions parents make regarding how and when they use digital media must balance the immediate needs of themselves and their children. The parent-child processes around digital media use are likely bidirectional, where child factors (ie, behavior problems) and digital media habits influence each other through a transactional process. For example, parents might use digital media to soothe or distract a fussy child [13], which reduces the frequency of enriching parent-child interactions [14-16] and contributes to continued behavioral difficulties [17,18], which in turn predict greater digital media exposure. The importance of understanding behavioral patterns that vary across contexts and families lies in the ability to identify the most salient causal pathways that can be used to develop individualized, tailored, and targeted intervention strategies [19]. Accurate and acceptable measurement of digital media use is necessary to elucidate the potentially unique mechanisms underlying digital media use and health behaviors.

Technological advances in passive mobile sensing have made it possible to access data already collected by mobile devices (ie, app use) to objectively measure digital media use [20,21]. By integrating intensive longitudinal data collected from multiple sources (ie, passive mobile screen time sensing, ecological momentary assessment [EMA], and accelerometry), we can reveal the process-oriented science that underlies the association between mobile phone use and other health

behaviors (ie, sleep and activity) and socioemotional health (ie, behavior problems) [22]. To our knowledge, no studies to date have leveraged a combination of EMA, accelerometry, and passive mobile sensing capabilities to specifically study digital media use among families with young children [11,23]. It is unclear whether this combination of methods can be successfully used to gather meaningful information about the digital media use of children and families.

### Objectives

Feasible and accurate measures of digital media use and context are a necessary first step toward understanding the dynamics of children's digital media use [10]. Therefore, this study aimed to evaluate the feasibility of a 2-wave intensive longitudinal data collection protocol designed to measure digital media use (ie, mobile phones and tablets), physical activity, sleep, sedentary behavior, and socioemotional context among caregiver-child dyads. The entire study protocol was administered twice (both instances being an average of 7 months apart) to examine the feasibility of recruiting and retaining a cohort willing to complete the protocol multiple times. Additional aims of this study were to describe the convergent validity of EMA measures and to describe preliminary agreement between caregiver self-reported digital media use and mobile phone use collected from passive mobile sensing.

## Methods

### Study Design

This pilot intensive longitudinal study aimed to recruit 100 caregiver-child dyads. The study used an observational case-crossover design [24], where dyads served as their own controls to assess the within- and between-day effects of immediate antecedents on a dependent variable measured multiple times throughout the day and week [24]. Caregivers and their 3- to 5-year-old children participated in a 30-day assessment protocol. The sample was invited back approximately 3-12 months after their initial assessment (depending on initial recruitment date) to complete a modified 14-day protocol.

### Ethical Considerations

The study protocol was approved by the institutional review board at the University of South Carolina in August 2020 (Pro00092634). Owing to COVID-19 protocol adjustments [25], consent was obtained remotely. Interested participants were directed to an informational website that described the study procedures and participant rights and protections and included a web-based consent form. Following the web-based consent process, eligible participants were contacted by a trained member of the research team by phone to verbally explain study procedures, answer questions, and confirm their desire to participate.

## Recruitment

Recruitment took place between September 2020 and September 2021. Recruitment and enrollment procedures are described elsewhere [26]. Briefly, we recruited a nonrandom volunteer sample by distributing flyers at daycare centers, pediatric clinics, and community centers. We also created boosted Facebook posts. In an effort to recruit a socioeconomically and racially diverse sample, we reached out directly to daycares serving low-income families and prioritized the enrollment of low-income families. Finally, we used a snowball recruitment strategy and compensated participants US \$10 if they referred another eligible family to participate in the study and if the referred family completed the protocol.

## Participants

Caregivers were eligible if they (1) were a primary caregiver of a child aged between 3 and 5 years (ie, preschool aged), (2) owned a smartphone device, and (3) were able to read and speak English. Exclusion criteria for children included a diagnosis of a severe developmental or physical disorder that would prevent ambulation.

## Study Protocol

### Overview

Study procedures are described in detail elsewhere [26]. Generally, the day-to-day operations of the study included participant communication, participant enrollment, and participant tracking. Daily operations were predominately conducted by 1 postdoctoral fellow, 1 doctoral student, and 1 staff member, with oversight from the principal investigator.

### First Wave of Data Collection

Briefly, after consenting, caregivers completed a baseline survey that contained measures of demographics, parental stress, screen time, psychological functioning, and child behavior. Caregivers completed EMAs for 7 days to measure child behavior problems, caregiver stress, and child screen time. Each day 4 signal-contingent EMAs were delivered to caregivers' mobile phones between 8:30 AM and 9 PM. The surveys expired after 2 hours and were delivered such that no overlapping time windows could exist. EMA assessments were conducted in the first week of the 30-day monitoring window. Additional details, including a full list of EMA questions, can be found elsewhere [26]. In addition, caregivers and children were asked to wear an Axivity AX3 accelerometer (Newcastle upon Tyne, United Kingdom) on their nondominant wrist for 24 hours per day (including while bathing and swimming) for 30 days to assess physical activity, sedentary behavior, and sleep. Nondominant wrist placement has been shown to improve compliance rates compared with waist placement [27]. Data were processed using GGIR (version 2.5.1) [28] in R (R Foundation for Statistical Computing). Mobile phone use over 30 days differed based on whether the caregiver used an iOS or an Android device. For participants with an Android device, we used Chronicle [29], an app designed specifically for passive screen time monitoring of Android devices. Caregivers were sent a link to download the app and asked to allow it to run on their phone during the study period and delete it after 30 days. Caregivers with an iOS phone were texted an automated reminder to send a screenshot

of their screen time use each day at 9 PM to a study-specific phone number. If a caregiver failed to send a correct screenshot during the 7 days of EMAs, study staff sent a personalized SMS text message within 24 hours. After the 7-day EMA period, study staff sent a personalized SMS text message if a caregiver failed to send a screenshot on 2 consecutive days (48 hours). Participants were compensated up to US \$180 for their participation in each wave of data collection (US \$360 in total; see [26] for more details). The number of reminders to send images was recorded as an indicator of feasibility. The use of technical support (participant use of research staff technical support as well as research staff requests from Chronicle support staff) was included as a screen time feasibility outcome. We attempted to conduct semistructured qualitative interviews with all participants who dropped out of the study.

### Second Wave of Data Collection

Approximately 3-12 months after their initial enrollment, families were invited to participate in the entire study protocol for a second time. The second wave was conducted to examine the feasibility of retaining a cohort sample and to assess the longer-term acceptability of the protocol in its entirety. Several protocol adjustments were made between waves 1 and 2 based on participant feedback from wave 1. Most notably, the monitoring period for EMA was increased from 7 to 14 days, and accelerometry and passive mobile sensing were shortened from 30 to 14 days to align with the EMA protocol. Logistically, shortening the accelerometer protocol allowed us to invite all participants to re-enroll in the second wave of data collection (only 50/100, 50% was initially proposed). In addition, *Qustodio* was used to passively monitor iOS devices (phones and tablets) and Amazon Kindle Fire tablets (a description of *Qustodio* is provided in the following section).

During the second wave of data collection, the study team began using *Qustodio*, a commercially available passive sensing app that allows a third party (ie, parents) to monitor the timing, duration, and content of digital media. Starting in October 2021, caregivers were invited to allow the research team to monitor their child's tablets (Kindle and iOS devices) using the media monitoring system, *Qustodio*. Starting in January 2022, caregivers with iOS devices were also asked to download *Qustodio* for passive mobile sensing instead of uploading screenshots, as was done in wave 1. To monitor devices with *Qustodio*, caregivers were sent a link to download the app. Participants with iOS devices were sent additional instructions to install a mobile configuration, as iOS did not allow third-party apps to monitor screen use directly. A research assistant worked with caregivers over the phone to complete the enrollment. Unlike *Chronicle*, each *Qustodio* enrollment required direct assistance from a member of the research team via call or text messaging (approximately 10 minutes per participant). Caregivers were instructed to delete the app and mobile configuration and unenroll after the study period (14 days). Enrollment of a child device was presented as optional for additional compensation, and the number of caregivers who agreed to enroll, number of successful enrollments, and number of days of data were collected as feasibility metrics.

## Primary Outcome Measures

The specific feasibility outcome measures for this pilot study were as follows:

- Recruitment rate was operationalized as the number of eligible families divided by the total number of families interested.
- Enrollment rate was defined as the number of participants enrolled divided by the number of participants approached to participate.
- Retention rate was defined as the number of participants who completed the study protocol (ie, engaged in each study task) divided by the total number of consenting and enrolled participants.
- Re-enrollment rate was defined as the number of participants who re-enrolled in the second wave of data collection divided by the number of participants who completed the initial wave of data collection.
- Mobile phone compliance was defined a priori as having 21 out of 30 (wave 1) and 10 out of 14 (wave 2) days of screen time data. For iOS users, this entailed texting at least 21 (wave 1) screen-use images and 10 (wave 2) screen-use images to research study staff or 10 out of 14 days of successfully monitored data using the Qustodio software. For Android users, this entailed having at least 21 (wave 1) and 10 (wave 2) days of successfully monitored data using the Chronicle software. Screen time compliance did not include participants who were unable to comply because of Technological issues (refer to the Technological Issues section). Participants were not offered an extension beyond 30 days to reach compliance.
- Accelerometer compliance was defined a priori as at least 21 of 30 days of valid accelerometer data for wave 1 and 10 of 14 days of valid accelerometer data for wave 2. A valid day of accelerometer data was defined as at least 16 hours of wear during a 24-hour period.
- EMA compliance was defined a priori as completing 21 out of 28 surveys for wave 1 and 39 out of 56 surveys for wave 2. We calculated the number of participants who reached strict compliance (wave 1: 21 surveys within a 7-day window and wave 2: 39 surveys within a 14-day window). Participants who were unable to meet strict compliance criteria within the designated window were offered an extension to reach compliance (not exceeding 30 days). The number of additional days needed to reach 21 surveys was also recorded.

## Secondary Outcome Measures

### Measurement Validity

In addition to testing protocol feasibility, we aimed to examine preliminary convergent and divergent validity of EMA measures of child behavior and parental stress with established measures of each construct, which were administered once at the beginning of the study period. We also aimed to examine the agreement between passive mobile phone sensing compared with a self-report questionnaire measure of mobile phone use, which is the current standard in many screen-use studies [30,31].

### Child Behavior

A full list of EMA items is described elsewhere [26]. Items were adapted from the Multidimensional Assessment of Preschool Disruptive Behavior scale [32] and assessed child tantrums, noncompliance, and aggression over the previous 2 hours. Behavior problems were aggregated at the person level and calculated as the count of EMA completions where caregivers endorsed a problem behavior divided by the total number of potential EMA completions when caregivers and children were together. EMA-measured behavior problems were compared with subscales on the Strengths and Difficulties Questionnaire (SDQ) [33], which was completed once during the baseline survey before the start of the EMA protocol.

### Caregiver Stress

Caregiver stress was assessed by EMA using 2 items used in previous EMA studies [34,35]. The items were “How stressed are you feeling right now?” and “How certain do you feel that you can deal with all the things that you have to do right now?” The items were rated on a 5-point and 4-point Likert scale, respectively, from “1 (not at all)” to “4 or 5 (extremely),” and the items were aggregated for each participant across all EMA time points. We then examined associations between EMA-measured stress and validated measures of stress as well as parental distress and psychological functioning, both of which were constructs hypothesized to be related to stress. Measures were completed once during the baseline survey before the start of the EMA protocol. The measured constructs included overall stress (assessed using the Perceived Stress Scale [36]) as well as caregiver distress (assessed using the Kansas Parental Satisfaction Scale [37] and Confusion, Hubbub, and Order Scale [CHAOS] [38]) and psychological functioning (assessed using the Center for Epidemiologic Studies-Depression [39] and the short-form State-Trait Anxiety Inventory [40]).

### Objective Phone Use (Passive Mobile Sensing)

We examined the preliminary agreement between passive mobile sensing screen time estimates compared with an existing widely used self-report questionnaire measure of technology use [41]. Specifically, caregivers answered the prompt “Thinking of an average weekday/weekend day (from when you wake up until you go to sleep), how much time do you spend using a smartphone as the primary activity?” Prompts for weekends and weekdays were assessed separately. For analyses, a composite measure was created that weighted weekend and weekday responses for an average daily use estimate. Notably, although this measure has shown good reliability, research on validity of self-report mobile phone use measures is largely absent from the literature. Thus, we refrain from using the term “convergent validity” when describing the associations between passive mobile sensing and self-report measures.

*Technological issues* with screen time monitoring were recorded and described. For iOS users, we monitored the number of days that required a personalized reminder prompt from the study team to send the image the following day. For Android (Chronicle), we tracked the number and type of technological complications as well as the number of participants impacted for the first 18 Android participants (between September 8,

2020, and July 7, 2021). The number of technological issues was only tracked for a subset of the study participants because of a change in help desk request procedures for Chronicle. Caregivers with an Android device were provided with instructions to enroll in Chronicle, an app designed specifically for passive screen time monitoring on Android devices [29].

### Protocol Modifications and Post Hoc Analyses

In this section, we describe protocol adjustments and modifications made over the course of the study. As these were not anticipated before the start of the study, we have elected to present them as post hoc and opportunistic analyses.

**Figure 1.** Decorated bands for child accelerometers.



In response to low initial compliance rates for child accelerometer wear, we examined options to increase compliance. In this opportunistic nonrandomized subanalysis, the first 30 children recruited into the study were sent an undecorated accelerometer band, after which wear compliance was preliminarily examined. In an effort to improve compliance, the next 75 children enrolled were sent decorated bands featuring iron-on decals of characters from Frozen, PAW Patrol, Marvel, or Blue's Clues, depending on the child's preferred characters (Figure 1).

### Statistical Analysis

All data were analyzed using SPSS (version 27; IBM Corp) and Stata SE (version 16.1; StataCorp). In this section, we present descriptive statistics of feasibility outcomes. Bayesian Pearson correlations (with a 95% credible interval) were used to assess the association between EMA problem behavior and SDQ subscales. Independent samples *t* tests (2-tailed) were used to examine differences between participants who completed the protocol and those who dropped out. Root mean square errors (RMSEs) were calculated to examine the agreement between parent-reported mobile phone use and objectively measured mobile phone use. As feasibility is the primary outcome of interest in this study, we examined the signal of effect difference using standard effect size estimates (ie, Cohen *d* and *r*) and minimal acceptable feasibility metrics in favor of significance

testing [42]. For post hoc tests of protocol additions, we conducted 2-tailed independent samples *t* tests to compare valid days of wear between participants who received decorated accelerometer bands and participants who received undecorated accelerometer bands.

### Sample Size

Given that this is a pilot study, no power analysis was required [43].

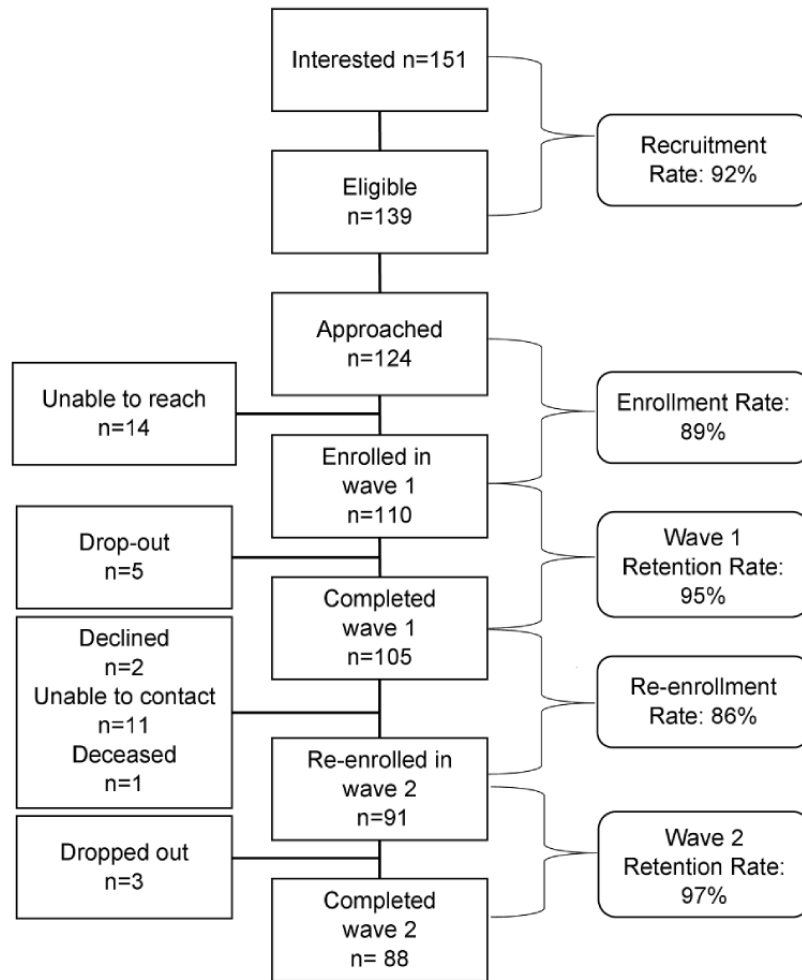
## Results

### Sample Characteristics

The flow of participants through the study is presented in Figure 2. Demographics of participants at wave 1 are presented in Table 1.



**Figure 2.** Flowchart showing recruitment and retention of study participants.



**Table 1.** Participant demographics (N=105).

Characteristics	Values
<b>Child demographics</b>	
Age (years), mean (SD)	4.4 (0.8)
<b>Race, n (%)</b>	
African American	32 (30)
White	66 (63)
Other	7 (7)
Hispanic ethnicity, n (%)	4 (4)
Female sex, n (%)	51 (49)
<b>Child device<sup>a</sup>, n (%)</b>	
None	16 (15)
Android	30 (29)
iOS	32 (30)
Kindle Fire	28 (27)
Other	9 (9)
Unknown or unable to verify	5 (5)
<b>Caregiver demographics</b>	
Age (years; range 22-78), mean (SD)	36.6 (7.8)
Female sex, n (%)	100 (95)
<b>Relationship with child, n (%)</b>	
Mother	95 (90)
Father	6 (6)
Grandparent	3 (3)
Other	1 (1)
<b>Phone type, n (%)</b>	
iPhone	65 (62)
Android	40 (38)
<b>Education, n (%)</b>	
Less than high school	1 (1)
High school	8 (8)
Some college or vocational training	25 (24)
2-year degree	13 (12)
4-year degree	31 (30)
Doctorate or professional degree	27 (26)
<b>Social assistance programs, n (%)</b>	
WIC <sup>b</sup>	25 (24)
SNAP <sup>c</sup>	28 (27)
Medicaid	48 (46)
<b>Income (US \$), n (%)</b>	
<20,000	14 (13)
20,000-40,000	19 (18)
40,000-60,000	28 (27)

Characteristics	Values
60,000-80,000	12 (11)
80,000-100,000	13 (12)
>100,000	19 (18)

<sup>a</sup>Child could have more than 1 device, so the percentages do not add up to 100%.

<sup>b</sup>WIC: Special Supplemental Nutrition Program for Women, Infants, and Children.

<sup>c</sup>SNAP: Supplemental Nutrition Assistance Program.

## Primary Outcomes

Primary recruitment, enrollment and re-enrollment, and retention rates are presented in Figure 2. Protocol compliance is presented in Table 2. All sample recruitment, retention, and enrollment

rates were >85%. Retention between the first and second waves of data collection was 86.7% (91/105). Participants were re-enrolled an average of 7.3 (SD 2.8) months after their initial enrollment. Outcomes are presented separately for wave 1 and wave 2.

**Table 2.** Compliance rates for primary outcome measures.

Outcome measure	Wave 1			Wave 2		
	Compliant/possible n	Participants meeting compliance criteria (%)	Number of valid observations, mean (SD; range)	Compliant/possible n	Compliant (%)	Number of valid observations, mean (SD; range)
<b>Accelerometry<sup>a</sup></b>						
Parent <sup>b</sup>	62/99	63	19.4 (7.5; 0-27)	72/84	86	13.4 (4.0; 0-22)
Child <sup>c</sup>	40/100	40	15.5 (8.2; 0-28)	57/85	67	10.8 (5.3; 0-20)
<b>Mobile phone<sup>a</sup></b>						
Android	38/40	95	32.0 (7.6; 7-53)	30/32	94	16.8 (4.8; 2-29)
iPhone	64/65	98	28.9 (2.9; 18-37)	54/56	96	16.3 (4.4; 0-30)
EMA <sup>d,e</sup>	105/105	100	26.2 (2.2; 21-32)	85/88	97	48.6 (6.9; 19-60)

<sup>a</sup>Wave 1 compliance was defined as 21 days of valid data provided; wave 2 compliance was defined as 9 days of valid data provided.

<sup>b</sup>Wave 1: lost device=1 and battery failure=5; wave 2: lost device=2 and battery failure=2.

<sup>c</sup>Wave 1: lost device=4 and battery failure=1; wave 2: lost device=3.

<sup>d</sup>EMA: ecological momentary assessment.

<sup>e</sup>Wave 1 compliance was defined as 21 prompts answered; wave 2 compliance was defined as 39 prompts answered.

## Wave 1

### Mobile Phone Monitoring Compliance

Screen time compliance is presented separately for Android and iOS devices, given the different data collection protocol procedures. A majority of both Android (38/40, 95%) and iOS (64/65, 98%) users met the study definition of compliance ( $\geq 21$  days of data). Notably, the remaining 5% (2/40) of Android participants did not meet the criteria because of technological issues with Chronicle, not study-related noncompliance. In all, 47.6% (50/105) of participants had over 30 days of data collection, attributable to the following reasons: (1) participants enrolling in screen time monitoring (Android) before receiving the activity watches and starting the protocol, (2) participants failing to delete the Chronicle app, or (3) participants sending additional screenshot images even after the end of their study observation period. The mean number of valid days of

observation data were 32.0 (SD 7.6) and 28.9 (SD 2.9) among Android and iOS users, respectively.

### Technological Issues

Technological issues emerged among both iOS and Android users. For iOS users, reminder prompts were sent on 32 occasions across 18 participants. Issues for iOS users were parents forgetting to send their nightly screenshots without an additional reminder from research study staff. For Chronicle, for the first 18 participants (September 2020 through July 7, 2021), there were 16 reports of technological issues. Issues for Android users were related to data processing problems with Chronicle, which were beyond the immediate control of the research staff (Table 3). Technological issues included (1) battery issues that occurred when the participant had battery optimization turned on their smartphones, (2) server malfunctioning, (3) large data sets causing data upload failures, and (4) data not uploading to the cloud from Chronicle.

**Table 3.** Technological issues.

	Number of incidents	Number of participants impacted/number of participants possible (%)
<b>Chronicle (Android)<sup>a</sup></b>		
Battery issues	6	6/18 (33)
Processing issues	11	11/18 (61)
<b>iOS (iPhone)</b>		
Number of reminder prompts sent	32	18/41 (44)

<sup>a</sup>Data were tracked between September 8, 2020, and July 7, 2021, across 18 Android participants.

### **Accelerometer Compliance**

Children wore the accelerometer for an average of 15.5 (SD 8.2) days during the 30-day monitoring period. A total of 4 children lost their device, and 1 device was returned with a battery malfunction. Among the children who returned a working device, 40% (40/100) met the study-specific compliance criteria of  $\geq 21$  days of data. Furthermore, 3 children returned the device but had 0 days of valid data. However, 91% (91/100) of the children had at least three days of valid data, a common criterion used in studies of children's physical activity [44].

Caregivers wore the accelerometer for an average of 19.4 (SD 7.5) days during the 30-day monitoring period. One caregiver lost the device, and 5 devices were returned with battery malfunctions. Among parents who returned a working device, 63% (62/99) met the study-specific compliance criteria of  $\geq 21$  days of data. In all, 96% (95/99) caregivers wore the device for at least three days.

### **EMA Compliance**

EMA compliance was high, with 100% (105/105) of caregivers completing at least 21 EMAs. EMAs were completed in an average of 2.8 (SD 10.24) minutes. In all, 95% of all EMA prompts were completed in  $\leq 5$  minutes. In all, 9.5% (10/105) of caregivers required additional opportunities to reach 21 completions. Of those who were provided with additional opportunities, caregivers were sent an average of 11 (SD 7.8) additional EMAs. Most participants (95/105, 90.5%) completed at least 21 EMAs, with no additional opportunities. Overall, caregivers completed an average of 26.2 (SD 2.2) total EMAs.

### **Dropouts**

There were minimal demographic differences between dropouts (5/110, 4.5%) and those who completed the protocol in terms of child age (Cohen  $d=-0.1$ ; 95% CI  $-1.1$  to  $0.9$ ), caregiver age (Cohen  $d=0.9$ ; 95% CI  $-0.1$  to  $1.9$ ), or employment status (Cohen  $d=0.3$ ; 95% CI  $-0.7$  to  $1.3$ ). The largest difference was observed for income (Cohen  $d=1.8$ ; 95% CI  $0.3$ - $2.4$ ), where participants who dropped out had lower incomes. One participant consented but never completed any study procedures. Of the 5 families who started the protocol and then dropped out, none of the caregivers agreed to participate in a semistructured qualitative interview about their experience. However, a review of text correspondence with project staff indicated that reasons for dropping out were lack of time (2/5, 40%) and no longer wanting to participate (1/5, 20%). The remaining 40% (2/5) of

participants stopped responding to study-related SMS text messages.

### **Wave 2**

#### **Mobile Phone Monitoring Compliance**

Screen time compliance is presented separately for Android and iOS devices, given the different data collection protocol procedures. A majority of both Android (30/32, 94%) and iOS (54/56, 96%) users met the study definition of compliance ( $\geq 10$  days of data). In all, 81% (71/88) of participants had over 14 days of data collection, attributable to the following reasons: (1) participants enrolling in screen time monitoring before receiving the activity watches and starting the protocol, (2) failing to delete the screen time monitoring app (Chronicle or Qustodio), (3) sending additional screenshot images after the end of the observation period, or (4) extending the study period to have extra EMA opportunities. The mean number of valid days of observation data were 16.8 (SD 4.8) and 16.3 (SD 4.4) among Android and iOS users, respectively.

#### **Accelerometer Compliance**

Children wore the accelerometer for an average of 10.8 (SD 5.3) days during the 14-day monitoring period. A total of 3 children lost their device. Among the children who returned a device, 67% (57/85) met the wave 2 study-specific compliance criteria of  $\geq 10$  days of data. Another 5% (4/85) of children returned the device but had 0 days of valid data. However, 87% (74/85) of the children had at least 3 days of valid data, a common criterion used in studies of children's physical activity [44]. Caregivers wore the accelerometer for an average of 13.4 (SD 4.0) days during the 14-day monitoring period. A total of 2 caregivers lost the device, and 2 devices were returned with a battery malfunction. One caregiver returned the device with 0 days of valid data. In all, 96% (81/84) of caregivers wore the device for at least three days, and 86% (72/84) of caregivers met the study-specific compliance criteria of  $\geq 10$  days.

#### **EMA Compliance**

EMA compliance was high, with 97% (85/88) of caregivers completing at least 39 EMAs. In all, 10% (9/88) of caregivers were provided with extra EMA opportunities to reach the compliance threshold. Of those who were provided with additional opportunities, caregivers were sent an average of 9 (SD 10.1) additional EMAs. Most participants (79/88, 90%) completed at least 39 EMAs with no additional opportunities. Overall, caregivers completed an average of 48.6 (SD 6.9) EMAs.

**Secondary Outcomes**

**Child Behavior**

Correlations are presented in Figure 3. Means and SDs are presented in Multimedia Appendix 1. Caregivers reported that their child’s behavior was “a little bit” to “a great deal” problematic in 31% (501/1614) of EMA instances. The most frequent problem behaviors were noncompliance (243/1614, 15.05%), tantrums (174/1614, 10.78%), and aggression (71/1614, 4.39%). For convergent validity, EMA aggression was correlated with SDQ subscales of conduct problems

( $r=0.389$ ; 95% CI 0.219-0.539), total difficulties ( $r=0.311$ ; 95% CI 0.140-0.479), peer problems ( $r=0.270$ ; 95% CI 0.092-0.439), and hyperactivity ( $r=0.195$ ; 95% CI 0.013-0.372). EMA noncompliance was correlated with conduct problems ( $r=0.334$ ; 95% CI 0.165-0.499) and total difficulties ( $r=0.247$ ; 95% CI 0.080-0.423). EMA tantrums were associated with peer problems ( $r=0.191$ ; 95% CI 0.008-0.371) and total difficulties ( $r=0.188$ ; 95% CI 0.000-0.361). The prosocial subscale, child age, and sex, were not strongly associated with aggression ( $r=-0.027$  to  $-0.142$ ), noncompliance ( $r=-0.165$  to  $-0.002$ ), or tantrums ( $r=-0.095$  to 0.015).

**Figure 3.** Correlations between Strengths and Difficulties subscales and ecological momentary assessment (EMA) behaviors of aggression, noncompliance, and tantrums.

Strengths and Difficulties subscale	Aggression	Noncompliance	Tantrum
Emotional problems	0.029	0.127	0.109
Conduct problems	0.389	0.334	0.162
Hyperactivity	0.195	0.120	0.096
Peer problems	0.270	0.130	0.191
Prosocial behavior	-0.173	-0.075	-0.076
Total difficulties	0.311	0.247	0.188
Biological sex	-0.142	-0.104	0.015
Child age	-0.066	-0.165	-0.028

Darker red indicates stronger positive correlation  
 Darker blue indicates stronger negative correlation

**Caregiver Stress**

Correlations are presented in Figure 4. Means and SDs are presented in Multimedia Appendix 1. EMA items regarding caregiver stress were compared with established measures of stress, parenting satisfaction, anxiety, depression, and household chaos. Caregivers who reported higher levels of stress on average across all EMA measures reported higher levels of overall stress on an established questionnaire ( $r=0.396$ ; 95% CI 0.222-0.542) as well as more anxiety ( $r=0.435$ ; 95% CI 0.277-0.581). Caregivers who reported higher average confidence in handling stress showed an inverse pattern; caregivers who had higher confidence showed less overall stress

( $r=0.543$ ; 95% CI  $-0.675$  to  $-0.410$ ), anxiety ( $r=-0.431$ ; 95% CI  $-0.580$  to  $-0.275$ ), and depression ( $r=-0.464$ ; 95% CI  $-0.613$  to  $-0.323$ ) and greater parenting satisfaction ( $r=0.286$ ; 95% CI 0.108-0.454). Caregivers who reported higher overall stress had more household disorganization ( $r=-0.374$ ; 95% CI  $-0.531$  to  $-0.212$ ), whereas caregivers who reported high levels of efficacy in being able to manage stress had lower household disorganization ( $r=0.419$ ; 95% CI 0.260-0.563). Average caregiver EMA-reported stress was less correlated with caregiver biological sex ( $r=0.043$ ; 95% CI  $-0.143$  to 0.230), age ( $r=-0.182$ ; 95% CI  $-0.360$  to  $-0.007$ ), or income ( $r=0.086$ ; 95% CI  $-0.107$  to 0.267). Similar patterns were observed for efficacy in managing stress.

**Figure 4.** Correlations between ecological momentary assessment (EMA) measures of stress and established measures of stress, parenting satisfaction, anxiety, depression, and household chaos. CHAOS: Confusion, Hubbub, and Order Scale - Lower scores indicate greater household disorganization.

	EMA Stress	EMA Handling
Stress	0.396	-0.543
Parenting satisfaction	-0.150	0.286
Depression	0.397	-0.464
Anxiety	0.427	-0.431
CHAOS	-0.374	0.419
Parent gender	0.043	-0.020
Parent age	-0.182	-0.003
Income	0.086	0.028

Darker red indicates stronger positive correlation  
 Darker blue indicates stronger negative correlation

### Caregiver Mobile Phone Screen Time

Average daily mobile phone use duration was 383.4 (SD 157.0) minutes for Android participants and 354.7 (SD 137.6) minutes for iOS users. For Android users, agreement with caregiver-reported mobile phone use was poor, with high rates of both underreporting and overreporting compared with self-reported mobile phone use. RMSE for Android was 157.1, indicating that, on average, caregivers misreported their mobile phone use by over 2.5 hours. For iOS users, RMSE was 81.4, indicating that caregivers misreported their mobile phone use by nearly 1.5 hours.

### Protocol Modifications and Post Hoc Analyses

#### Accelerometer Bands

A total of 6 children were excluded because of lost or malfunctioning devices, and 56 children were included in the post hoc analysis. Of 26 nondecorated bands, 11 children wore the Axivity for ≥21 days (42% compliance). Of 30 decorated bands, 18 children wore the Axivity for ≥21 days (60% compliance). Children who received decorated bands had significantly more days of valid wear (mean 19.83, SD 5.56) than those who received undecorated bands (mean 14.38, SD 10.17;  $t_{37.46} = -2.44$ ;  $P = .02$ ; Cohen  $d = 8.02$ ). Given these midstudy results, the bands continued to be decorated for the remaining participants.

#### Child Device Enrollment

During the second wave of data collection (starting October 2021), 90 caregiver-child dyads were invited to monitor their children’s devices. Of the 90 caregiver-child dyads invited, 72 (80%) indicated that their child had their own device, 96% (69/72) of which were compatible with Chronicle or Qustodio. Although caregivers from 92% (66/72) of the families agreed to enroll a child device, 88% (63/72) both agreed to enroll and had a compatible device. Ultimately, 86% (54/63) of families were successfully enrolled. Unsuccessful enrollment was

because of the following reasons: technological issues installing the Qustodio app (3/72, 4%), difficulties scheduling a time to enroll the child’s device (3/72, 4%), child device lost (1/72, 1%), caregiver decided not to enroll the child’s device because of lack of use (1/72, 1%), caregiver being unable to access the child’s device (eg, “locked out”; 1/72, 1%). Overall, an average of 16.8 (SD 8.2) days of data per person were captured on child devices using Qustodio and Chronicle.

## Discussion

### Principal Findings

This is the first cohort study to use passive mobile sensing in conjunction with accelerometry and EMA measures with caregiver-child dyads. This preliminary work demonstrates that measures and protocols to collect these multiple streams of data are feasible. The convergent validity between EMA-measured child behavior and caregiver stress shows preliminary validity for using EMA to measure stress and behavior. The generally high levels of compliance and retention of the study sample are encouraging signs that these multiple overlapping methods of intensive longitudinal data can be completed in a longitudinal cohort study. The lack of agreement between self-reported and objectively measured mobile phone use highlights the need for additional objective and low-burden methods of measuring mobile phone use.

Feasibility outcomes revealed that participants were generally compliant with passive mobile sensing but that technology issues resulted in some data loss. It is notable that this technology is still relatively new, and ideally, as this technology becomes more refined, fewer technological issues will arise. Caregivers were generally willing to allow us to monitor their own device, as evidenced by their enrollment in study procedures and high rates of screen time monitoring compliance. Among families where children had their own device, 92% (66/72) of caregivers were willing to let us monitor their

children's device. Although we were only able to collect data on a portion of those devices because of the novelty of this technology of our monitoring software, continued advancement and expanding ease of use of monitoring software would likely increase this number, which bodes well for future studies aiming to examine child digital media use.

Initially, we observed poor accelerometer compliance among children, but this improved following the implementation of a protocol to decorate the accelerometer bands. Notably, the low compliance is likely a function of our strict valid day criteria (ie, 16 hours vs the more commonly used 10 hours) [44] coupled with our relatively high metric of compliance (21 days). Our 21-day compliance is higher than most accelerometer studies, which commonly set compliance at 3 days [44] and use a minimum of 10 hours of wear to determine a valid day. Although not a direct comparison, if we use a criterion of 3 days of valid data (with a 16-hour valid day cutoff), our study shows similar rates of compliance (17% noncompliance) as other studies that use wrist-worn accelerometers (22.7% noncompliance) [45].

This study also demonstrated the feasibility of deploying EMA measures for at least 7 days. Not only did participants complete most prompts but nearly all were also compliant per study protocol definitions and were willing to complete additional days of measurement in subsequent waves of data collection (ie, 14 days). It is worth noting that although 100% (105/105) of the sample were considered compliant using criteria specified a priori, some families needed additional days to accumulate the requisite number of completed EMA prompts. Similar to previous EMA studies [46], this flexibility was built into the protocol to retain participants from diverse and low-income backgrounds who might not be able to meet study requirements within the given time span, given time constraints, resources, acute stressors, or complex life situations. Furthermore, the EMA measures of child behavior and stress used in this study showed preliminary evidence of convergent validity with established validated measures.

Overall, across all study measures, participants reported minimal burden and expressed willingness to continue to participate in future waves. Finally, we were able to recruit and retain a sample to complete measures twice, both instances being an average of 7 months apart, indicating the potential for such methods to be used in cohort studies across development.

### Comparison With Prior Work

Other studies have used EMA among racially and ethnically diverse households [46] but not in conjunction with passive mobile sensing and coupled with both caregiver and child accelerometry. The confluence of these data is necessary to understand how and why screen habits develop and ultimately influence children's sleep, activity, and socioemotional health. Combining objective digital media use data with microtemporal EMA data will allow us to identify systems of "Granger causality," which are systems where one behavior (eg, digital media use) predicts future behavior (ie, tantrums). Ensuring that these measures and procedures are both feasible and accurate is a necessary first step toward this goal. The links among behaviors likely vary in direction and magnitude among

different individuals. Characterizing these links can guide advanced intervention methods (eg, just-in-time adaptive interventions and continuous tuning interventions) that aim to maximize change in multiple health behaviors.

### Limitations and Next Steps

This study was designed to overcome current measurement limitations regarding mobile phone use. Although our nonrandom sampling approach does represent a threat to external validity (ie, families who volunteer for a study may differ from the general population of families), our successful recruitment and retention of an economically and racially diverse sample speaks to the generalizability of our study findings. In terms of monitoring digital media use, the use of passive sensing represents a significant step forward. However, several key limitations still exist with regard to passive mobile sensing as a method of measuring digital media use. A significant limitation to this work, and to the field at large, is the inability to distinguish who is using a given device (ie, parent, child, or sibling). Although we attempted to ask caregivers when their child was using their smartphone using EMA, these methods were only able to validate to the standard of self-report, which as discussed earlier is inherently flawed. Furthermore, passive mobile sensing is only able to detect whether a screen is on and not whether it is being actively viewed, which in theory could overestimate digital media use. Although some research has attempted to distinguish who is watching a screen based on facial recognition using cameras [47], this work is preliminary and is not currently available for portable digital media. However, although passive mobile sensing is not able to capture all screen use (such as television), this technology represents a significant advancement in the field. Given the recent dramatic increases in children's use of digital media specifically, and the relatively outdated research on screen time more generally, there is unique value in measuring and investigating digital media use specifically. Research on digital media use specifically (unique from overall screen-use duration) is necessary to understand the unique causes and consequences of this new technology relative to health outcomes. However, low-burden, accurate, and objective measures of digital media use are necessary to advance the science underlying the dynamics and causal factors and outcomes around digital media use.

Although we did not observe any dropouts because of privacy concerns in this study, the issue of data privacy is worth noting. Chronicle was built specifically as a research tool, designed to be Health Insurance Portability and Accountability Act-compliant, with participant protections in mind. Among other features, the dashboard requires secure credentials for separate users, and data are deidentified and not linked to any IP addresses or phone numbers [29]. Participants were informed that Chronicle tracks mobile device use in terms of whether the phone is on or off, which apps are running, and what time of the day the apps are used. Chronicle does not collect personal information, phone contacts, content of any SMS text messages or emails, information on what websites are visited, or other specifics such as which videos are watched on YouTube. In contrast, third-party apps such as Qustodio have capacities that may not be useful or desirable to researchers (eg, the ability to log keystrokes and website blocking capability). Researchers

using passive mobile sensing apps should be aware of data security (how data are being stored and transmitted) as well as data ownership (to whom do the data belong once they are collected). Although these threats are not unique to research using mobile technologies, there is a need to develop consent processes that actively engage individuals in their own privacy decision-making as much as possible [48,49].

## Conclusions

Overall, this study demonstrates that it is largely feasible to collect intensive longitudinal data on objective digital media use, simultaneously with accelerometry and EMA, from an economically and racially diverse sample of families with preschool-aged children. Although this study represents an initial improvement in objective measurement of digital media, additional measurement work is needed to advance the field to understand digital media use in the context of interpersonal dynamics and health.

## Acknowledgments

This study is supported by the National Institute of General Medical Sciences of the National Institutes of Health (P20GM130420).

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Means and SDs of study measures (n=105).

[DOCX File, 14 KB - [formative\\_v6i9e40572\\_app1.docx](#) ]

## References

1. Tomopoulos S, Dreyer BP, Valdez P, Flynn V, Foley G, Berkule SB, et al. Media content and externalizing behaviors in Latino toddlers. *Ambul Pediatr* 2007 May;7(3):232-238. [doi: [10.1016/j.ambp.2007.02.004](#)] [Medline: [17512884](#)]
2. Hinkley T, Verbestel V, Ahrens W, Lissner L, Molnár D, Moreno LA, IDEFICS Consortium. Early childhood electronic media use as a predictor of poorer well-being: a prospective cohort study. *JAMA Pediatr* 2014 May 01;168(5):485-492. [doi: [10.1001/jamapediatrics.2014.94](#)] [Medline: [24639016](#)]
3. Stiglic N, Viner RM. Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews. *BMJ Open* 2019 Jan 03;9(1):e023191 [FREE Full text] [doi: [10.1136/bmjopen-2018-023191](#)] [Medline: [30606703](#)]
4. Carter B, Rees P, Hale L, Bhattacharjee D, Paradkar MS. Association between portable screen-based media device access or use and sleep outcomes: a systematic review and meta-analysis. *JAMA Pediatr* 2016 Dec 01;170(12):1202-1208 [FREE Full text] [doi: [10.1001/jamapediatrics.2016.2341](#)] [Medline: [27802500](#)]
5. World Health Organization. WHO Guidelines on Physical Activity, Sedentary Behaviour and Sleep for Children Under 5 Years of Age. Geneva: World Health Organization; 2018.
6. Chaput J, Colley RC, Aubert S, Carson V, Janssen I, Roberts KC, et al. Proportion of preschool-aged children meeting the Canadian 24-hour movement guidelines and associations with adiposity: results from the Canadian health measures survey. *BMC Public Health* 2017 Nov 20;17(Suppl 5):829 [FREE Full text] [doi: [10.1186/s12889-017-4854-y](#)] [Medline: [29219075](#)]
7. Kabali H, Irigoyen M, Nunez-Davis R, Budacki JG, Mohanty SH, Leister KP, et al. Exposure and use of mobile media devices by young children. *Pediatrics* 2015 Dec;136(6):1044-1050. [doi: [10.1542/peds.2015-2151](#)] [Medline: [26527548](#)]
8. Rideout V. The common sense census: Media use by kids age zero to eight. San Francisco, CA: Common Sense Media 2017:263-283. [doi: [10.4324/9781315002095-8](#)]
9. Vandewater EA, Lee S. Measuring children's media use in the digital age: issues and challenges. *Am Behav Sci* 2009 Apr 01;52(8):1152-1176 [FREE Full text] [doi: [10.1177/0002764209331539](#)] [Medline: [19763246](#)]
10. Browne DT, May SS, Colucci L, Hurst-Della Pietra P, Christakis D, Asamoah T, MIST Working Group. From screen time to the digital level of analysis: a scoping review of measures for digital media use in children and adolescents. *BMJ Open* 2021 May 19;11(5):e046367 [FREE Full text] [doi: [10.1136/bmjopen-2020-046367](#)] [Medline: [34011597](#)]
11. Byrne R, Terranova CO, Trost SG. Measurement of screen time among young children aged 0-6 years: a systematic review. *Obes Rev* 2021 Aug 07;22(8):e13260 [FREE Full text] [doi: [10.1111/obr.13260](#)] [Medline: [33960616](#)]
12. Chen C, Chen S, Wen P, Snow CE. Are screen devices soothing children or soothing parents? Investigating the relationships among children's exposure to different types of screen media, parental efficacy and home literacy practices. *Comput Human Behav* 2020 Nov;112:106462. [doi: [10.1016/j.chb.2020.106462](#)]
13. Radesky JS, Peacock-Chambers E, Zuckerman B, Silverstein M. Use of mobile technology to calm upset children: associations with social-emotional development. *JAMA Pediatr* 2016 Apr 01;170(4):397-399. [doi: [10.1001/jamapediatrics.2015.4260](#)] [Medline: [26928293](#)]



14. Machida S, Taylor AR, Kim J. The role of maternal beliefs in predicting home learning activities in head start families. *Family Relations* 2002 Apr;51(2):176-184. [doi: [10.1111/j.1741-3729.2002.00176.x](https://doi.org/10.1111/j.1741-3729.2002.00176.x)]
15. Nikken P. Parents' instrumental use of media in childrearing: relationships with confidence in parenting, and health and conduct problems in children. *J Child Fam Stud* 2018 Nov 16;28(2):531-546. [doi: [10.1007/s10826-018-1281-3](https://doi.org/10.1007/s10826-018-1281-3)]
16. Giallo R, Treyvaud K, Cooklin A, Wade C. Mothers' and fathers' involvement in home activities with their children: psychosocial factors and the role of parental self-efficacy. *Early Child Development Care* 2013 Apr;183(3-4):343-359. [doi: [10.1080/03004430.2012.711587](https://doi.org/10.1080/03004430.2012.711587)]
17. Munzer TG, Miller AL, Wang Y, Kaciroti N, Radesky JS. Tablets, toddlers and tantrums: the immediate effects of tablet device play. *Acta Paediatr* 2021 Jan 18;110(1):255-256 [FREE Full text] [doi: [10.1111/apa.15509](https://doi.org/10.1111/apa.15509)] [Medline: [32735726](https://pubmed.ncbi.nlm.nih.gov/32735726/)]
18. Hiniker A, Suh H, Cao S, Kientz J. Screen time tantrums: how families manage screen media experiences for toddlers and preschoolers. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2016; San Jose California USA. [doi: [10.1145/2858036.2858278](https://doi.org/10.1145/2858036.2858278)]
19. Glenn AL, Lochman JE, Dishion T, Powell NP, Boxmeyer C, Kassing F, et al. Toward tailored interventions: sympathetic and parasympathetic functioning predicts responses to an intervention for conduct problems delivered in two formats. *Prev Sci* 2019 Jan 8;20(1):30-40 [FREE Full text] [doi: [10.1007/s1121-017-0859-0](https://doi.org/10.1007/s1121-017-0859-0)] [Medline: [29308549](https://pubmed.ncbi.nlm.nih.gov/29308549/)]
20. Ryding FC, Kuss DJ. Passive objective measures in the assessment of problematic smartphone use: a systematic review. *Addict Behav Rep* 2020 Jun;11:100257 [FREE Full text] [doi: [10.1016/j.abrep.2020.100257](https://doi.org/10.1016/j.abrep.2020.100257)] [Medline: [32467846](https://pubmed.ncbi.nlm.nih.gov/32467846/)]
21. Radesky J, Weeks H, Ball R, Schaller A, Yeo S, Durnez J, et al. Young children's use of smartphones and tablets. *Pediatrics* 2020 Jul;146(1):e20193518 [FREE Full text] [doi: [10.1542/peds.2019-3518](https://doi.org/10.1542/peds.2019-3518)] [Medline: [32482771](https://pubmed.ncbi.nlm.nih.gov/32482771/)]
22. Dunton GF. Sustaining health-protective behaviors such as physical activity and healthy eating. *JAMA* 2018 Aug 21;320(7):639-640 [FREE Full text] [doi: [10.1001/jama.2018.6621](https://doi.org/10.1001/jama.2018.6621)] [Medline: [29852046](https://pubmed.ncbi.nlm.nih.gov/29852046/)]
23. Jones A, Armstrong B, Weaver RG, Parker H, von Klinggraeff L, Beets MW. Identifying effective intervention strategies to reduce children's screen time: a systematic review and meta-analysis. *Int J Behav Nutr Phys Act* 2021 Sep 16;18(1):126 [FREE Full text] [doi: [10.1186/s12966-021-01189-6](https://doi.org/10.1186/s12966-021-01189-6)] [Medline: [34530867](https://pubmed.ncbi.nlm.nih.gov/34530867/)]
24. Maclure M, Mittleman AM. Should we use a case-crossover design? *Annu Rev Public Health* 2000 May;21(1):193-221. [doi: [10.1146/annurev.publhealth.21.1.193](https://doi.org/10.1146/annurev.publhealth.21.1.193)] [Medline: [10884952](https://pubmed.ncbi.nlm.nih.gov/10884952/)]
25. Turner-McGrievy G, Halliday TM, Moore JB. COVID-19 messed up my research: insights from physical activity and nutrition translational research. *Transl J Am Coll Sports Med* 2021 Aug 13;6(4):e000169. [doi: [10.1249/tjx.0000000000000169](https://doi.org/10.1249/tjx.0000000000000169)] [Medline: [34693004](https://pubmed.ncbi.nlm.nih.gov/34693004/)]
26. Reesor-Oyer L, Parker H, Burkart S, Smith M, Dugger R, von Klinggraeff L, et al. Measuring Microtemporal Processes Underlying Preschoolers' Screen Use and Behavioral Health: Protocol for the Tots and Tech Study. *JMIR Res Protoc* 2022 Sep 28;11(9):e36240. [doi: [10.2196/36240](https://doi.org/10.2196/36240)]
27. Fairclough SJ, Noonan R, Rowlands AV, Van Hees V, Knowles Z, Boddy LM. Wear compliance and activity in children wearing wrist- and hip-mounted accelerometers. *Med Sci Sports Exerc* 2016 Mar;48(2):245-253. [doi: [10.1249/MSS.0000000000000771](https://doi.org/10.1249/MSS.0000000000000771)] [Medline: [26375253](https://pubmed.ncbi.nlm.nih.gov/26375253/)]
28. Migueles J, Rowlands A, Huber F, Sabia S, van HV. GGIR: a research community-driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. *J Measurement Physical Behav* 2019;2(3):188-196. [doi: [10.1123/jmpb.2018-0063](https://doi.org/10.1123/jmpb.2018-0063)]
29. Barr R, Kirkorian H, Radesky J, Coyne S, Nichols D, Blanchfield O, et al. Beyond screen time: a synergistic approach to a more comprehensive assessment of family media exposure during early childhood. *Front Psychol* 2020 Jul 10;11:1283 [FREE Full text] [doi: [10.3389/fpsyg.2020.01283](https://doi.org/10.3389/fpsyg.2020.01283)] [Medline: [32754078](https://pubmed.ncbi.nlm.nih.gov/32754078/)]
30. Griffioen N, van Rooij M, Lichtwarck-Aschoff A, Granic I. Toward improved methods in social media research. *Technol Mind Behav* 2020 Jun 17;1(1). [doi: [10.1037/tmb0000005](https://doi.org/10.1037/tmb0000005)]
31. Parry D, Davidson B, Sewall C, Fisher J, Mieczkowski H, Quintana D. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nat Hum Behav* 2021 Nov;5(11):1535-1547. [doi: [10.1038/s41562-021-01117-5](https://doi.org/10.1038/s41562-021-01117-5)] [Medline: [34002052](https://pubmed.ncbi.nlm.nih.gov/34002052/)]
32. Wakschlag LS, Briggs-Gowan MJ, Choi SW, Nichols SR, Kestler J, Burns JL, et al. Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. *J Am Acad Child Adolesc Psychiatry* 2014 Jan;53(1):82-96.e3 [FREE Full text] [doi: [10.1016/j.jaac.2013.10.011](https://doi.org/10.1016/j.jaac.2013.10.011)] [Medline: [24342388](https://pubmed.ncbi.nlm.nih.gov/24342388/)]
33. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry* 2001 Nov;40(11):1337-1345. [doi: [10.1097/00004583-200111000-00015](https://doi.org/10.1097/00004583-200111000-00015)] [Medline: [11699809](https://pubmed.ncbi.nlm.nih.gov/11699809/)]
34. O'Connor SG, Habre R, Bastain TM, Toledo-Corral CM, Gilliland FD, Eckel SP, et al. Within-subject effects of environmental and social stressors on pre- and post-partum obesity-related biobehavioral responses in low-income Hispanic women: protocol of an intensive longitudinal study. *BMC Public Health* 2019 Mar 28;19(1):253 [FREE Full text] [doi: [10.1186/s12889-019-6583-x](https://doi.org/10.1186/s12889-019-6583-x)] [Medline: [30819155](https://pubmed.ncbi.nlm.nih.gov/30819155/)]
35. Berge J, Tate A, Trofholz A, Fertig AR, Miner M, Crow S, et al. Momentary parental stress and food-related parenting practices. *Pediatrics* 2017 Dec;140(6):e20172295 [FREE Full text] [doi: [10.1542/peds.2017-2295](https://doi.org/10.1542/peds.2017-2295)] [Medline: [29167378](https://pubmed.ncbi.nlm.nih.gov/29167378/)]

36. Cohen S. Perceived stress in a probability sample of the United States. In: *The Social Psychology of Health*. Thousand Oaks, California: SAGE Publications; 1988.
37. James DE, Schumm WR, Kennedy CE, Grigsby CC, Shectman KL, Nichols CW. Characteristics of the Kansas parental satisfaction scale among two samples of married parents. *Psychol Rep* 2016 Sep 01;57(1):163-169. [doi: [10.2466/pr0.1985.57.1.163](https://doi.org/10.2466/pr0.1985.57.1.163)]
38. Matheny AP, Wachs TD, Ludwig JL, Phillips K. Bringing order out of chaos: psychometric characteristics of the confusion, hubbub, and order scale. *J Applied Developmental Psychol* 1995 Jul;16(3):429-444. [doi: [10.1016/0193-3973\(95\)90028-4](https://doi.org/10.1016/0193-3973(95)90028-4)]
39. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D. *Am J Prevent Med* 1994 Mar;10(2):77-84. [doi: [10.1016/s0749-3797\(18\)30622-6](https://doi.org/10.1016/s0749-3797(18)30622-6)]
40. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *Br J Clin Psychol* 1992 Sep 12;31(3):301-306. [doi: [10.1111/j.2044-8260.1992.tb00997.x](https://doi.org/10.1111/j.2044-8260.1992.tb00997.x)] [Medline: [1393159](https://pubmed.ncbi.nlm.nih.gov/1393159/)]
41. Vizcaino M, Buman M, DesRoches CT, Wharton C. Reliability of a new measure to assess modern screen time in adults. *BMC Public Health* 2019 Oct 28;19(1):1386 [FREE Full text] [doi: [10.1186/s12889-019-7745-6](https://doi.org/10.1186/s12889-019-7745-6)] [Medline: [31660931](https://pubmed.ncbi.nlm.nih.gov/31660931/)]
42. Orsmond GI, Cohn ES. The distinctive features of a feasibility study: objectives and guiding questions. *OTJR (Thorofare N J)* 2015 Jul 06;35(3):169-177. [doi: [10.1177/1539449215578649](https://doi.org/10.1177/1539449215578649)] [Medline: [26594739](https://pubmed.ncbi.nlm.nih.gov/26594739/)]
43. 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](https://doi.org/10.1136/bmj.i5239)] [Medline: [27777223](https://pubmed.ncbi.nlm.nih.gov/27777223/)]
44. Cain K, Sallis J, Conway T, Van Dyck D, Calhoun L. Using accelerometers in youth physical activity studies: a review of methods. *J Phys Act Health* 2013 Mar;10(3):437-450 [FREE Full text] [doi: [10.1123/jpah.10.3.437](https://doi.org/10.1123/jpah.10.3.437)] [Medline: [23620392](https://pubmed.ncbi.nlm.nih.gov/23620392/)]
45. Howie EK, Straker LM. Rates of attrition, non-compliance and missingness in randomized controlled trials of child physical activity interventions using accelerometers: a brief methodological review. *J Sci Med Sport* 2016 Oct;19(10):830-836. [doi: [10.1016/j.jsams.2015.12.520](https://doi.org/10.1016/j.jsams.2015.12.520)] [Medline: [26874648](https://pubmed.ncbi.nlm.nih.gov/26874648/)]
46. Trofholz A, Tate A, Janowiec M, Fertig A, Loth K, de Brito JN, et al. Ecological momentary assessment of weight-related behaviors in the home environment of children from low-income and racially and ethnically diverse households: development and usability study. *JMIR Res Protoc* 2021 Dec 01;10(12):e30525 [FREE Full text] [doi: [10.2196/30525](https://doi.org/10.2196/30525)] [Medline: [34855612](https://pubmed.ncbi.nlm.nih.gov/34855612/)]
47. Vadathya AK, Musaad S, Beltran A, Perez O, Meister L, Baranowski T, et al. An objective system for quantitative assessment of television viewing among children (family level assessment of screen use in the home-television): system development study. *JMIR Pediatr Parent* 2022 Mar 24;5(1):e33569 [FREE Full text] [doi: [10.2196/33569](https://doi.org/10.2196/33569)] [Medline: [35323113](https://pubmed.ncbi.nlm.nih.gov/35323113/)]
48. Trifan A, Oliveira M, Oliveira JL. Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. *JMIR Mhealth Uhealth* 2019 Aug 23;7(8):e12649 [FREE Full text] [doi: [10.2196/12649](https://doi.org/10.2196/12649)] [Medline: [31444874](https://pubmed.ncbi.nlm.nih.gov/31444874/)]
49. 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](https://doi.org/10.2196/mhealth.4538)] [Medline: [26474545](https://pubmed.ncbi.nlm.nih.gov/26474545/)]

## Abbreviations

**CHAOS:** Confusion, Hubbub, and Order Scale

**EMA:** ecological momentary assessment

**RMSE:** root mean square error

**SDQ:** Strengths and Difficulties Questionnaire

*Edited by A Mavragani; submitted 27.06.22; peer-reviewed by A Yee, J de Brito; comments to author 27.07.22; revised version received 16.08.22; accepted 17.08.22; published 29.09.22.*

*Please cite as:*

*Parker H, Burkart S, Reesor-Oyer L, Smith MT, Dugger R, von Klinggraeff L, Weaver RG, Beets MW, Armstrong B  
Feasibility of Measuring Screen Time, Activity, and Context Among Families With Preschoolers: Intensive Longitudinal Pilot Study  
JMIR Form Res 2022;6(9):e40572*

*URL: <https://formative.jmir.org/2022/9/e40572>*

*doi: [10.2196/40572](https://doi.org/10.2196/40572)*

*PMID: [36173677](https://pubmed.ncbi.nlm.nih.gov/36173677/)*

©Hannah Parker, Sarah Burkart, Layton Reesor-Oyer, Michal T Smith, Roddrick Dugger, Lauren von Klinggraeff, R Glenn Weaver, Michael W Beets, Bridget Armstrong. Originally published in JMIR Formative Research (<https://formative.jmir.org>),

29.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Extracurricular Humanism in Medicine Initiative and Medical Student Wellness: Retrospective Study

Elizabeth Diane Auckley<sup>1\*</sup>, BA; Jeff Barbee<sup>2\*</sup>, MA, DMA; Nicole Verbeck<sup>2\*</sup>, MPH; Tracie McCambridge<sup>3\*</sup>, MA; Linda Stone<sup>1\*</sup>, MA, MD; Jennifer Garvin<sup>4</sup>, MBA, RHIA, PhD

<sup>1</sup>The Ohio State University College of Medicine, Columbus, OH, United States

<sup>2</sup>Office of Curriculum and Scholarship, The Ohio State University College of Medicine, Columbus, OH, United States

<sup>3</sup>Department of Medical Education, The Ohio State University College of Medicine, Columbus, OH, United States

<sup>4</sup>Health Information Management and Systems Divisions, Ohio State University, Columbus, OH, United States

\* these authors contributed equally

**Corresponding Author:**

Elizabeth Diane Auckley, BA

The Ohio State University College of Medicine

370 W 9th Ave

Columbus, OH, 43210

United States

Phone: 1 4402926035

Email: [elizabeth.auckley@osumc.edu](mailto:elizabeth.auckley@osumc.edu)

## Abstract

**Background:** Humanism in Medicine Initiative (HIMI), an extracurricular program at Ohio State University College of Medicine (OSUCOM) with 27 subgroups, fosters the humanities. Stress and burnout among first- and second-year medical students are prevalent across the United States. Solutions for stress among first- and second-year medical students have been proposed, but no gold standard exists. The relationship of humanism with stress and burnout has yet to be described in the literature.

**Objective:** This study investigates the relationship between participation in the HIMI and stress, burnout, and academic success among first- and second-year medical students.

**Methods:** First- and second-year medical students enrolled at OSUCOM between August 2018 and August 2019 were recruited. Attendance in the HIMI and membership records were used to measure their participation. Curricular examination scores and those on Step 1 of United States Medical Licensing Examination (USMLE) were used to measure academic success. Stress and burnout were measured using the Maslach Burnout Inventory and the Perceived Stress Scale.

**Results:** In total, 412 students were enrolled with 362 (87%) students participating in HIMI. Those with high participation were more often Black, Asian, female, or with a humanities undergraduate major compared to the overall study population. There were significant relationships between Gold Humanism Honors Society (GHHS) induction and participation of first- and second-year medical students in service- ( $\chi^2_1=5.8$ ,  $P<.05$ ) or leadership-focused ( $\chi^2_1=19.3$ ,  $P<.001$ ) HIMI groups. Medium levels of participation in the HIMI were associated with significantly higher stress. Performance on the Step 1 USMLE was not significantly associated with participation levels in the HIMI (low=233.7 vs high=238.0;  $P=.10$ ).

**Conclusions:** The HIMI is an extracurricular program vastly utilized by first- and second-year medical students at OSUCOM and did not impact Step 1 USMLE scores. Medium participation in the HIMI was associated with higher stress, and service- and leadership-focused HIMI participation was associated with a higher level of induction to the GHHS. This study identifies areas for future studies to understand the relationship of the HIMI with stress and academic success.

(*JMIR Form Res* 2022;6(9):e37252) doi:[10.2196/37252](https://doi.org/10.2196/37252)

**KEYWORDS**

humanism; extracurricular; stress; burnout; medical student; student; academic success; wellness

## Introduction

Stress and burnout are well known to the majority of medical students in the United States [1-5]. Stress was first described as the “nonspecific response of the body to any demand made upon it” [6] but has since been recognized as a potentially negative state of being, where the body is threatened by conditions that create imbalance and endanger homeostasis [7]. Burnout, initially described as excessive burdens of energy, strength, or resources in work environments, resulting in malaise, fatigue, cynicism, frustration, and inefficacy [8], is now well understood as a syndrome of emotional exhaustion, depersonalization, and decreased personal achievement due to human interaction at the workplace [2,9]. Stress and burnout often embody a cyclical, perpetuating relationship [2,10] that preclinical medical students are uniquely vulnerable to experiencing [1,10]. In the already hypercompetitive medical school environment with frequent high-stakes formal assessments, students cite unanticipated informal verbal quizzing, feeling useless, and depersonalized preclinical medical education as specific sources of stress and burnout in addition to sources from life outside of school [1,11]. Both psychological conditions are associated with psychological morbidity, anxiety and depression, alcohol and drug abuse, suicidal ideation, and suicide in medical students [1,12,13].

Despite being recognized as a prevalent issue among medical students for at least 4 decades [14,15], no gold standard of stress and burnout reduction for this population has been described [12]. The focus of improvement has shifted from treating impacted students individually to changing the environment of preclinical medical school to be less stressful overall [15]. However, stress and burnout solutions for medical students remain under investigated, with few studies including empirical data for specific interventions [12]. Intervention in medical school is critical, as these psychological conditions in physicians have been found to originate from early-career training in medical school [14]. Further, physicians with resilience, an important phenomenon to fight stress and burnout, are more prepared to create successful patient-physician relationships because they are better able to provide hope and empathy to patients [16].

While stress and burnout are increasingly becoming an important consideration of medical education reforms, humanism has also gained the attention of medical educators in recent decades. The Arnold P. Gold Foundation defines a humanistic health care practitioner as one with integrity, excellence, collaboration and compassion, altruism, respect and resilience, empathy, and service [17]. The Arnold P. Gold Foundation recognizes humanism in medical students through the national honors society induction for the Gold Humanism Honors Society (GHHS). The GHHS selects members based on peer nominations and individual applications for those with strong nominations. Induction into GHHS at OSCOM requires strong peer nominations during the third year of medical school, followed by an invitation to complete an essay-based application. GHHS members select new inductees after rigorous review of deidentified applications and peer nominations. GHHS induction is intended to be impartial. Local chapters of GHHS

are limited to inducting a maximum of 15% of each class. Humanism is often considered the foundation for professionalism, which is defined as a way of acting to meet the expectations of the workplace. Professionalism only becomes authentic when humanism, the morals and values that guide ones obligation to serve others especially when in need, is fostered [18]. Calls to protect innate humanism in students [19], foster growth of humanism in medical school culture [18,20,21], and add humanism to the top 4 goals of health care [22] are supported by staff and students alike [23], but specific initiatives and measurement of humanism has been difficult to describe [18-20,24]. Further, the impact of humanities and culture of humanism on preclinical medical students’ stress and burnout has not been previously reported, based on our review of the literature.

This study investigates the impact of first- and second-year medical students’ engagement with HIMI on their stress, burnout, and academic success. HIMI, a portion of the Linda C. Stone MD Program for Humanism and the Arts in Medicine, is a voluntary extracurricular program at the Ohio State University College of Medicine (OSUCOM), which has been developed and is led by students with the goal of fostering a culture of humanism among first- and second-year medical students. Begun in 2009, it now embodies 27 unique student organizations that allow students to engage in the arts and humanities with their peers. Many isolated aspects of HIMI have been supported as general stress- and burnout-reducing activities for students in the literature, including animal therapy [25-27], music [5], peer mentor programs [16,28,29], emotional expression with peers [29-32], student retreats [29], visual arts [33], and being student-led [13].

## Methods

### Ethical Approval

This study was approved by the institutional review board at the Ohio State University (IRB# 2020B0173) and approved for waived consent.

### Recruitment

The setting for this study was a single-center retrospective study conducted at OSUCOM (Columbus, Ohio), a large, Midwestern medical school. The medical school curriculum is structured as two years of lectures and small group-based classes with bimonthly clinical experiences. The third year involves clinical rotations with weekly case discussions, and the fourth year requires specialty specific rotations, an emergency medicine clinical rotation, and an advanced elective. Within this curriculum, Humanism is one option for fourth-year elective courses, and a Humanism in Medicine extracurricular group exists to supplement medical student experiences on an optional basis. This study involved medical students enrolled at OSUCOM from August 2018 to August 2019. All first- and second-year [34] medical students were eligible and enrolled in the study unless they opted out. Students were given the option to terminate participation by opting out, through an email notification. Student demographic information was obtained from internal admissions records for those enrolled.

## Participation and Leadership in the HIMI

First- and second-year medical student participation in the HIMI is defined as attending any event hosted by a student group in the HIMI, organizing an event, leading a group or event, or nominating a peer for a nomination-based group during August 2018 to July 2019. Participation was measured by internal HIMI records of attendance, membership, and leadership. Any amount of participation in a single HIMI group qualified for participation for that group, and the total number of unique HIMI groups participated in was measured for each student. In 2018-2019, HIMI encompassed 27 unique student groups, categorized as art-focused (eg, Dance in Medicine and Photography/Film), service-focused (eg, Life Support Group and Peer Mentoring), or leadership-focused (eg, Medicine and the Arts Board and Servant Leadership). Student leaders, who had been nominated or recognized by the HIMI as a student organizer of the group, were also recorded.

## Academic Outcomes

Academic success was measured on the basis of the Step 1 United States Medical Licensing Examination (USMLE) score and student Objective Structured Clinical Examination (OSCE) scores. The OSCEs have been administered at most medical schools for more than 4 decades [35] with the goal of measuring communication skills and clinical tasks through directed interactions with several standardized patients [36]. OSCE scores are reported as percentages (0%-100%), and the last OSCE score for the 2018-2019 school year was used for analysis. Step 1 is the first of 3 examinations in the USMLE series. This is a nationally standardized examination required for medical school graduation and licensing that examines student understanding of basic biomedical principles and clinical applications. Step 1 is administered after the completion of didactic coursework and is taken after the second year at OSUCOM [37]. Scores are reported numerically out of 300 and correspond to percentiles nationally [38]. OSCE and Step 1 USMLE scores were obtained from the Office of Curriculum and Scholarship at OSUCOM.

## Stress and Burnout

Stress and burnout were measured using the Maslach Burnout Inventory (MBI), Perceived Stress Scale (PSS), Perceived Cohesion Scale (PCS) [39], and quality of life (QOL) measure. These surveys were administered independently by OSUCOM in a voluntary survey to students in spring 2018. The MBI is based on a 7-point scale of frequency from "never" to "daily" responses to 22 statements organized in 3 sections corresponding to the 3 parts of burnout: emotional exhaustion (9 statements), depersonalization (5 statements), and reduced personal accomplishment (8 statements) [9]. The MBI's items are derived from the 1996 Maslach Burnout Inventory Manual. The MBI has been validated as a measure of burnout in preclinical medical students [40]. The PSS is a measure of how one's current life circumstances are interpreted as stressful [41]. A 10-item form of the Likert-scale-based self-report PSS instrument was used, and scores were reported between 10 and 50, with lower scores indicating lower stress. The PSS is validated for populations with at least junior high school education and has been reported to have an internal consistency of 0.85 [42]. The PCS measures

sense of belonging and morale within a population and records a response from 0="strongly disagree" to 10="strongly agree" to 6 statements, 3 pertaining to sense of belonging and 3 to morale [43]. The population noted in the statements was "medical school class" and "medical school" in this survey. A revised, Bollen and Hoyle form of the Likert-scale self-report PCS [43] was used, which reported scores between 1 and 60, and lower scores indicated less perceived cohesion. QOL was measured using a linear analogue self-assessment, which scores QOL as 1-5, with 1="as bad as it can be" and 5="as good as it can be"; this QOL measure has been validated in a wide set of populations [44].

## GHHS Induction

Induction into the GHHS was recorded from internal records of the GHHS. The GHHS was established by the Arnold P. Gold Foundation in 2002 and recognizes students who embody humanistic qualities such as empathy and integrity and serve as role models and advocates for humanism in medicine. The GHHS has been added as a qualification on the Electronic Residency Application Service for students to identify themselves as members, suggesting its importance [45].

## Statistical Analysis

To understand the potential relationship between participating in an HIMI group (ie, arts, leadership-, or service-focused organization) and induction into the GHHS, second-year students, who were eligible to be inducted into GHSS the following year, were grouped dichotomously as having participated in at least one activity, or not at all. A chi-square goodness-of-fit test was conducted for each category of the HIMI group to determine if the proportions of students who participated in at least one group and inducted into the GHHS was similar to those of students who did not participate in any HIMI groups.

Examining the HIMI program and wellness constructs such as burnout and perceived stress, student participation was demarcated into 3 levels of participation: low (0-1 groups), medium (2-3 groups), and high (4 or more groups). This grouping was performed to better differentiate potential effects and attempt to obtain evenly balanced groups. ANOVA was performed to identify any significant differences. All assumptions were checked to ensure this was the appropriate analysis technique.

When studying the perceptions of belonging and morale from a cohort and school viewpoint, the limited data required the participation variable to be collapsed to a dichotomous grouping. The low participation group attended 1 or 0 HIMI groups, while the high participation group attended  $\geq 2$  groups. Owing to failure of the assumption of normality, the nonparametric Kruskal-Wallis test was applied. These survey instruments were optional for students to complete and required additional consent for research.

Dichotomous grouping into low and high group participation groups was also used to test for significant relationships with the Step 1 USMLE scores of second-year students. An independent samples *t* test was performed for this analysis. This

analytical approach was continued to examine the Professionalism OSCEs during the last second-year evaluation.

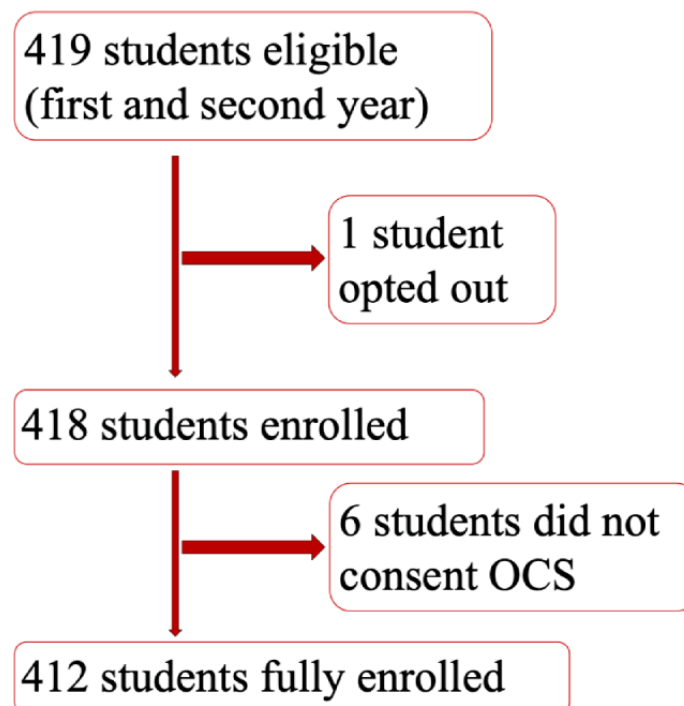
The final relationship studied was between participation in the HIMI and the Interpersonal-Communication Skills OSCE scores of first-year students. This OSCE was conducted during the Endocrine-Reproduction block. These data allowed for the use of the trichotomous grouping of low, middle, and high participation levels. ANOVA was performed for this analysis with no assumption violations. All statistical analyses were performed using RStudio (version 1.3.959; RStudio Team).

## Results

### Recruitment and Demographics

In total, 419 first- and second-year students were enrolled at OSUCOM during the study period, and 412 met inclusion criteria for analysis (Figure 1). Regarding race and ethnicity, the study population comprised 48.8% White (n=201), 8.5% Black (n=35), and 19.9% Asian (n=82) students, and 22.8% of participants did not respond (n=94), and regarding gender, the study population comprised 45.4% females (n=188) and 41.7% males (n=172), and 12.8% of participants did not respond (n=52). Most participants did not have a humanities undergraduate major (n=338, 82.0%) as opposed to 8.4% (n=35) of participants who had a humanities major, and 9.6% (n=39) of participants did not respond (Table 1).

**Figure 1.** Recruitment of participants. There were 419 first- and second-year students enrolled in Ohio State University College of Medicine (OSUCOM) in 2018-2019. After an email invitation to opt out of the study, 1 student opted out of the study. Six students did not provide consent to the Office of Curriculum and Scholarship (OCS) for the use of demographic and academic data. The analysis included the remaining 412 students. Surveys were optional for students and some data were incomplete.



**Table 1.** Demographic information regarding participants in the Humanism in Medicine Initiative (HIMI) between 2018-2019 (N=412).

	Low participation in the HIMI (n=180), n (%)	Medium participation in the HIMI (n=147), n (%)	High participation in the HIMI (n=85), n (%)	Total, n (%)
<b>Race</b>				
Asian	30 (16.7)	30 (20.4)	22 (25.8)	82 (19.9)
Black	19 (10.6)	8 (5.4)	8 (9.4)	35 (8.5)
White	86 (47.8)	80 (54.4)	35 (41.2)	201 (48.8)
No response	45 (24.9)	29 (19.8)	20 (23.6)	94 (22.8)
<b>Gender</b>				
Female	59 (32.7)	75 (50.7)	54 (63.5)	188 (45.4)
Male	90 (50.0)	61 (41.5)	21 (24.7)	172 (41.7)
No response	31 (17.3)	11 (7.8)	10 (11.8)	52 (12.8)
<b>Undergraduate major</b>				
Humanities	11 (6.1)	16 (10.8)	8 (9.4)	35 (8.4)
Other	142 (78.8)	125 (85.0)	71 (83.5)	338 (82.0)
No response	27 (14.8)	6 (4.2)	6 (7.1)	39 (9.6)

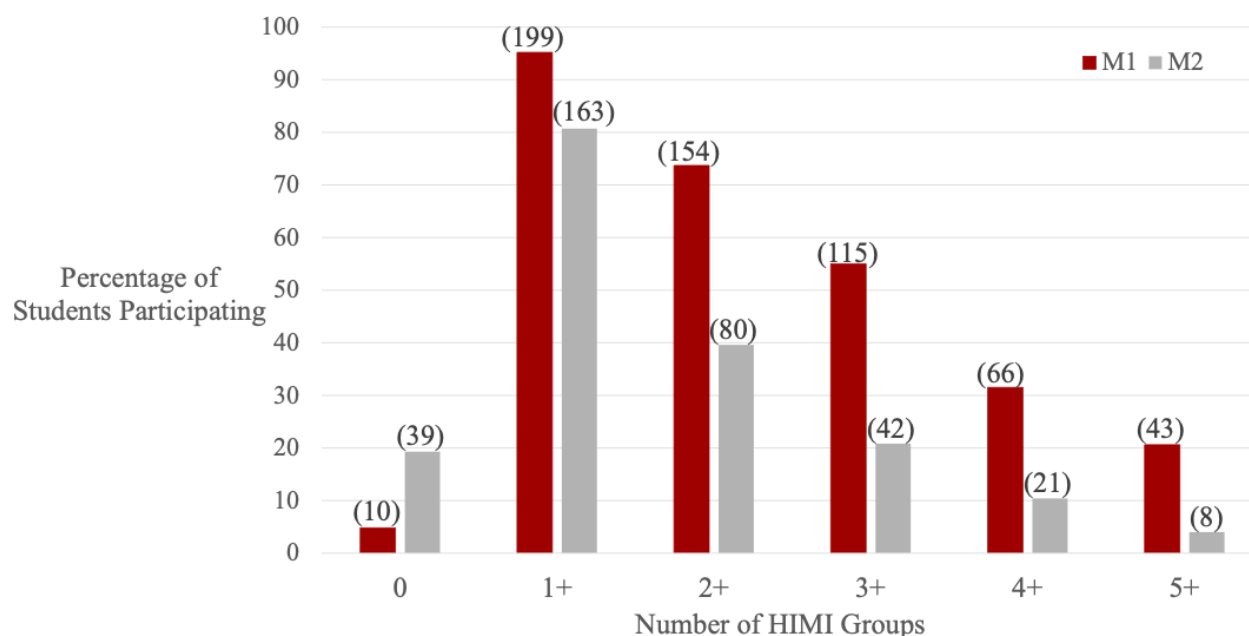
Race, gender, and undergraduate major of the participants were obtained from the Office of Curriculum and Scholarship internal records, which were optionally self-reported by students. No response indicates the percentage of students who did not opt to self-report their information.

### Participation in the HIMI

First- and second-year students vastly participated in HIMI student groups. The HIMI participation rate was 95% (n=199

students) among first-year students and 81% (n=163) among second-year students, accounting for a total participation rate of 87% (n=362). First-year students averaged participating in 3.0 unique student groups, and second-year in 1.6 unique student groups (Figure 2). Of the HIMI groups, Peer Mentoring (n=333), Somali Health Initiative for Nutrition Education (n=62), Visual Arts (n=49), M1 Fall Retreat (n=43), and Dance in Medicine (n=39) were those most participated in.

**Figure 2.** Medical student participation in the Humanism in Medicine Initiative (HIMI) groups, displayed by minimum participation levels. Each HIMI group includes all students who participated in the minimum amount stated or more. Medical student participation in HIMI groups was defined as attending any event hosted by a student group in the HIMI, organizing an event, leading a group or event, and nominating a peer for a nomination-based group between August 2018 and July 2019 and collected from internal HIMI records.



Student participation was categorized as low (0-1 groups), medium (2-3 groups), or high ( $\geq 4$  groups). When compared

with the total study population, high participation was observed to a larger extent among Black or Asian (Black: n=35, 8.5% vs



n=8, 9.4%; Asian: n=82, 19.9% vs 22, 25.8%), among female participants (n=188, 45.4% vs n=54, 63.5%) and those who have majored in humanities during their undergraduate studies (n=35, 8.4% vs n=8, 9.4%).

### Academic Outcome Results

Step 1 USMLE scores were analyzed for second-year students. HIMI participation had no significant relationship with the OSCE outcomes. When comparing low and high participation, the high participation group averaged higher scores (238.0, SD 14.6) than the low participation group (233.7, SD 19.3) ( $P=.10$ ), though the difference was not significant.

### Stress and Burnout

Perceived stress was found to be a significant factor by participation level ( $F_{2,193}=3.85$ ;  $P<.05$ ;  $\eta_p^2=0.038$ ). A post hoc analysis was conducted given the significance of the omnibus test. The Tukey honestly significant difference multiple comparison test was used to examine all pairwise relationships. A significant difference was found between the students with medium participation (18.5, SD 6.6) and students with low participation (15.7, SD 5.58;  $P=.02$ ). This finding indicates that medium participation in the HIMI was significantly associated with higher levels of perceived stress than those with low participation in the HIMI. Participation in the HIMI was not significantly associated with burnout, sense of belonging, or QOL.

### GHHS Induction

GHHS induction was significantly associated with participation in a HIMI service ( $\chi^2_1=5.8$ ,  $P<.05$ ) and leadership focused group ( $\chi^2_1=19.3$ ,  $P<.001$ ).

## Discussion

### Principal Findings

This study investigated the short-term impact of first- and second-year students' participation in the HIMI on their stress, burnout, and academic success. We found that the HIMI is utilized by the majority of first- and second-year students (95% vs 81% participation, respectively), and those who participated in  $\geq 4$  HIMI groups were more often female, Black, Asian, or had a humanities undergraduate major compared to the total study population.

Further, students who participated in a service- or leadership-focused group were significantly more likely to be inducted into the GHHS than those who did not participate in these groups.

Medium participation (defined as participation in 2-3 HIMI groups), compared with low participation (0-1 groups), was associated with significantly higher stress. Medical student stress and burnout have been a concern in recent decades with no gold-standard solution described. The association of medium HIMI participation with high stress warrants future study at our institution, with more medical students, and at other institutions with similar programs. Perhaps students with higher baseline stress seek out HIMI for reduction, but it is difficult to determine

the cause of stress, as a complex multidimensional variable, and the results of this study cannot infer causation. Furthermore, participation in the HIMI had no significant impact on Step 1 USMLE scores.

### Comparison With Prior Work

Humanities content during undergraduate medical education has also been a topic of debate for decades. Most recent concerns seek to determine how to balance the need for humanistic training and diverting students' time from rigorous basic science study to accomplish this. Moreover, the success of humanities content as part of the curriculum versus as an extracurricular program has similarly been debated. Our results suggest that students highly utilize the HIMI as a voluntary opportunity to engage with the humanities. This study found first-year students to engage with HIMI to a greater degree than second-year students, possibly owing to the increased demands on second-year students to study for the Step 1 USMLE. HIMI membership was also more commonly observed among students who are female, Black, Asian and had a humanities background than the overall medical school population. One possible reason for this result could be that students with an identity underrepresented in medicine (female, Black, or Asian students and those with a humanities education) sought out the HIMI as a community of support.

The association of HIMI leadership and service engagement with GHHS membership suggests that participation in the HIMI makes students more competitive for GHHS membership, possibly either by characterizing time spent on valued activities or identification of intrinsic interest in humanities. While the GHHS and HIMI appear to share several values, the GHHS selection process requires strong peer nominations as the first step for consideration. HIMI involvement may therefore add holistic review of applicants.

However, students' time spent in the HIMI did not significantly impact Step 1 USMLE scores in our study, and these findings may be evidence for including humanities content in the medical students' schedule without concern for detracting from basic or clinical science education.

### Limitations

This study was completed at a single medical school in an academic institution with a small sample size of students, compared to the population of medical students in the United States. The results of the study are therefore limited in generalizability within the United States and internationally. The study constituted a secondary data analysis and was limited in analysis by existing data. Further variables could be analyzed in future, experimentally designed studies. Participation data were collected internally by students and may be limited by program design. Further, participation was classified binarily and thus did not discriminate difference in volume of participation within each group. Some student group participation may be underrepresented in the study owing to these limitations. Student demographic and some outcome data were optional to report, which may have led to bias in our results. The analysis had a cross-sectional design and did not account for longitudinal outcomes of participation in the HIMI.

The study utilized solely quantitative data, which may limit the ability to capture the full impact of humanism on medical students [46]. Finally, the study was observational by design.

### Conclusions

The HIMI is an extracurricular program vastly utilized by first- and second-year medical students at OSUCOM and did not impact Step 1 USMLE scores. Medium participation in the

HIMI was associated with higher stress, and service- and leadership-focused HIMI participation was associated with a higher level of induction to the GHHS. This study identified areas for future studies to understand the association of participation in the HIMI with stress, academic success, potentially important activities for social support and resilience in medical school.

### Acknowledgments

The authors wish to thank OSUCOM, the Office of Curriculum and Scholarship at Ohio State University College of Medicine, and the Humanism in Medicine Research Team.

### Conflicts of Interest

None declared.

### References

- Moffat KJ, McConnachie A, Ross S, Morrison JM. First year medical student stress and coping in a problem-based learning medical curriculum. *Med Educ* 2004 May;38(5):482-491. [doi: [10.1046/j.1365-2929.2004.01814.x](https://doi.org/10.1046/j.1365-2929.2004.01814.x)] [Medline: [15107082](https://pubmed.ncbi.nlm.nih.gov/15107082/)]
- Fares J, Al Tabosh H, Saadeddin Z, El Mouhayyar C, Aridi H. Stress, Burnout and Coping Strategies in Preclinical Medical Students. *N Am J Med Sci* 2016 Feb;8(2):75-81 [FREE Full text] [doi: [10.4103/1947-2714.177299](https://doi.org/10.4103/1947-2714.177299)] [Medline: [27042604](https://pubmed.ncbi.nlm.nih.gov/27042604/)]
- Mazurkiewicz R, Korenstein D, Fallar R, Ripp J. The prevalence and correlations of medical student burnout in the pre-clinical years: a cross-sectional study. *Psychol Health Med* 2012 Mar;17(2):188-195. [doi: [10.1080/13548506.2011.597770](https://doi.org/10.1080/13548506.2011.597770)] [Medline: [21781020](https://pubmed.ncbi.nlm.nih.gov/21781020/)]
- Dyrbye LN, West CP, Satele D, Boone S, Tan L, Sloan J, et al. Burnout Among U.S. Medical Students, Residents, and Early Career Physicians Relative to the General U.S. Population. *Academic Medicine* 2014;89(3):443-451. [doi: [10.1097/acm.000000000000134](https://doi.org/10.1097/acm.000000000000134)]
- Fares J, Saadeddin Z, Al Tabosh H, Aridi H, El Mouhayyar C, Koleilat MK, et al. Extracurricular activities associated with stress and burnout in preclinical medical students. *J Epidemiol Glob Health* 2016 Sep;6(3):177-185 [FREE Full text] [doi: [10.1016/j.jegh.2015.10.003](https://doi.org/10.1016/j.jegh.2015.10.003)] [Medline: [26644345](https://pubmed.ncbi.nlm.nih.gov/26644345/)]
- Selye H. Stress without Distress. In: *Psychopathology of Human Adaptation*. New York, NY: Springer; 1976:137-146.
- Nechita F, Nechita D, Pîrlog MC, Rogoveanu I. Stress in medical students. *Rom J Morphol Embryol* 2014;55(3 Suppl):1263-1266 [FREE Full text] [Medline: [25607418](https://pubmed.ncbi.nlm.nih.gov/25607418/)]
- Reith T. Burnout in United States Healthcare Professionals: A Narrative Review. *Cureus* 2018 Dec 04;10(12):e3681 [FREE Full text] [doi: [10.7759/cureus.3681](https://doi.org/10.7759/cureus.3681)] [Medline: [30761233](https://pubmed.ncbi.nlm.nih.gov/30761233/)]
- Maslach C, Schaufeli W, Marek T. *Professional Burnout Recent Developments in Theory and Research*. London: Routledge; 1993.
- Vyas K, Stratton T, Soares N. Sources of medical student stress. *Educ Health* 2017;30(3):232. [doi: [10.4103/efh.efh\\_54\\_16](https://doi.org/10.4103/efh.efh_54_16)]
- Firth J. Levels and sources of stress in medical students. *Br Med J (Clin Res Ed)* 1986 May 03;292(6529):1177-1180 [FREE Full text] [doi: [10.1136/bmj.292.6529.1177](https://doi.org/10.1136/bmj.292.6529.1177)] [Medline: [3085772](https://pubmed.ncbi.nlm.nih.gov/3085772/)]
- Shapiro SL, Shapiro DE, Schwartz GER. Stress management in medical education: a review of the literature. *Acad Med* 2000 Jul;75(7):748-759. [doi: [10.1097/00001888-200007000-00023](https://doi.org/10.1097/00001888-200007000-00023)] [Medline: [10926029](https://pubmed.ncbi.nlm.nih.gov/10926029/)]
- Laster J, Geary C, Sierpina VS, Kreitzer MJ. A student initiative for reducing stress in medical students: Stress-Less-Fest. *Explore (NY)* 2012 Sep;8(5):314-316. [doi: [10.1016/j.explore.2012.06.012](https://doi.org/10.1016/j.explore.2012.06.012)] [Medline: [22938752](https://pubmed.ncbi.nlm.nih.gov/22938752/)]
- Wolf T, Balson P, Faucett J, Randall H. A retrospective study of attitude change during medical education. *Med Educ* 1989 Jan;23(1):19-23. [doi: [10.1111/j.1365-2923.1989.tb00807.x](https://doi.org/10.1111/j.1365-2923.1989.tb00807.x)] [Medline: [2927336](https://pubmed.ncbi.nlm.nih.gov/2927336/)]
- Huebner LA, Royer JA, Moore J. The assessment and remediation of dysfunctional stress in medical school. *J Med Educ* 1981 Jul;56(7):547-558. [doi: [10.1097/00001888-198107000-00002](https://doi.org/10.1097/00001888-198107000-00002)] [Medline: [7241552](https://pubmed.ncbi.nlm.nih.gov/7241552/)]
- Dunn LB, Iglewicz A, Moutier C. A conceptual model of medical student well-being: promoting resilience and preventing burnout. *Acad Psychiatry* 2008 Feb 01;32(1):44-53. [doi: [10.1176/appi.ap.32.1.44](https://doi.org/10.1176/appi.ap.32.1.44)] [Medline: [18270280](https://pubmed.ncbi.nlm.nih.gov/18270280/)]
- Gaufberg E. Mapping the Landscape, Journeying Together. *Academic Medicine* 2017;92(12):1671-1673. [doi: [10.1097/acm.0000000000001987](https://doi.org/10.1097/acm.0000000000001987)]
- Cohen JJ. Viewpoint: Linking Professionalism to Humanism: What It Means, Why It Matters. *Academic Medicine* 2007;82(11):1029-1032. [doi: [10.1097/01.acm.0000285307.17430.74](https://doi.org/10.1097/01.acm.0000285307.17430.74)]
- Martimianakis MA, Michalec B, Lam J, Cartmill C, Taylor JS, Hafferty FW. Humanism, the Hidden Curriculum, and Educational Reform. *Academic Medicine* 2015;90:S5-S13. [doi: [10.1097/acm.0000000000000894](https://doi.org/10.1097/acm.0000000000000894)]

20. Fins JJ, Pohl B, Doukas DJ. In Praise of the Humanities in Academic Medicine. *Camb Q Healthc Ethics* 2013 Aug 12;22(4):355-364. [doi: [10.1017/s0963180113000200](https://doi.org/10.1017/s0963180113000200)]
21. Bifulco M, Pisanti S. Integrating Medical Humanities into medical school training. *EMBO Rep* 2019 Dec 05;20(12):e48830 [FREE Full text] [doi: [10.15252/embr.201948830](https://doi.org/10.15252/embr.201948830)] [Medline: [31663234](https://pubmed.ncbi.nlm.nih.gov/31663234/)]
22. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med* 2014 Nov 10;12(6):573-576 [FREE Full text] [doi: [10.1370/afm.1713](https://doi.org/10.1370/afm.1713)] [Medline: [25384822](https://pubmed.ncbi.nlm.nih.gov/25384822/)]
23. Pories SE, Piawah S, Abel GA, Mullangi S, Doyle J, Katz JT. What is the Role of the Arts in Medical Education and Patient Care? A Survey-based Qualitative Study. *J Med Humanit* 2018 Dec 4;39(4):431-445. [doi: [10.1007/s10912-018-9530-4](https://doi.org/10.1007/s10912-018-9530-4)] [Medline: [30076508](https://pubmed.ncbi.nlm.nih.gov/30076508/)]
24. Bowman D. On being 'good' in the medical humanities. *Med Humanit* 2016 Mar;42(1):1-2. [doi: [10.1136/medhum-2016-010890](https://doi.org/10.1136/medhum-2016-010890)] [Medline: [26933199](https://pubmed.ncbi.nlm.nih.gov/26933199/)]
25. Binfet J, Passmore H, Cebry A, Struik K, McKay C. Reducing university students' stress through a drop-in canine-therapy program. *J Ment Health* 2018 Jun 21;27(3):197-204. [doi: [10.1080/09638237.2017.1417551](https://doi.org/10.1080/09638237.2017.1417551)] [Medline: [29265945](https://pubmed.ncbi.nlm.nih.gov/29265945/)]
26. Ward-Griffin E, Klaiber P, Collins HK, Owens RL, Coren S, Chen FS. Petting away pre-exam stress: The effect of therapy dog sessions on student well-being. *Stress Health* 2018 Aug 12;34(3):468-473. [doi: [10.1002/smi.2804](https://doi.org/10.1002/smi.2804)] [Medline: [29528189](https://pubmed.ncbi.nlm.nih.gov/29528189/)]
27. Wood E, Ohlsen S, Thompson J, Hulin J, Knowles L. The feasibility of brief dog-assisted therapy on university students stress levels: the PAwS study. *J Ment Health* 2018 Jun 06;27(3):263-268. [doi: [10.1080/09638237.2017.1385737](https://doi.org/10.1080/09638237.2017.1385737)] [Medline: [28984144](https://pubmed.ncbi.nlm.nih.gov/28984144/)]
28. Mouret GML. Stress in a graduate medical degree. *Med J Aust* 2002 Jul;177(S1). [doi: [10.5694/j.1326-5377.2002.tb04652.x](https://doi.org/10.5694/j.1326-5377.2002.tb04652.x)]
29. Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. *Med Educ* 2016 Jan 23;50(1):132-149. [doi: [10.1111/medu.12927](https://doi.org/10.1111/medu.12927)] [Medline: [26695473](https://pubmed.ncbi.nlm.nih.gov/26695473/)]
30. Branch W, Pels RJ, Lawrence RS, Arky R. Becoming a Doctor -- Critical-Incident Reports from Third-Year Medical Students. *N Engl J Med* 1993 Oct 07;329(15):1130-1132. [doi: [10.1056/nejm199310073291518](https://doi.org/10.1056/nejm199310073291518)]
31. Maslach C. Burned-out. *Can J Psychiatr Nurs* 1979;20(6):5-9. [Medline: [260904](https://pubmed.ncbi.nlm.nih.gov/260904/)]
32. Pololi L, Frankel RM. Small-group teaching emphasizing reflection can positively influence medical students' values. *Acad Med* 2001 Dec;76(12):1172; author reply 1172-1173. [doi: [10.1097/00001888-200112000-00002](https://doi.org/10.1097/00001888-200112000-00002)] [Medline: [11739036](https://pubmed.ncbi.nlm.nih.gov/11739036/)]
33. Haidet P, Jarecke J, Adams NE, Stuckey HL, Green MJ, Shapiro D, et al. A guiding framework to maximise the power of the arts in medical education: a systematic review and metasynthesis. *Med Educ* 2016 Mar 19;50(3):320-331. [doi: [10.1111/medu.12925](https://doi.org/10.1111/medu.12925)] [Medline: [26896017](https://pubmed.ncbi.nlm.nih.gov/26896017/)]
34. Vanderbilt A, Feldman M, Wood I. Assessment in undergraduate medical education: a review of course exams. *Med Educ Online* 2013 Mar 06;18(1):1-5 [FREE Full text] [doi: [10.3402/meo.v18i0.20438](https://doi.org/10.3402/meo.v18i0.20438)] [Medline: [23469935](https://pubmed.ncbi.nlm.nih.gov/23469935/)]
35. Harden R, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13(1):41-54. [doi: [10.1111/j.1365-2923.1979.tb00918.x](https://doi.org/10.1111/j.1365-2923.1979.tb00918.x)]
36. Cömert M, Zill JM, Christalle E, Dirmaier J, Härter M, Scholl I. Assessing Communication Skills of Medical Students in Objective Structured Clinical Examinations (OSCE)--A Systematic Review of Rating Scales. *PLoS One* 2016 Mar 31;11(3):e0152717 [FREE Full text] [doi: [10.1371/journal.pone.0152717](https://doi.org/10.1371/journal.pone.0152717)] [Medline: [27031506](https://pubmed.ncbi.nlm.nih.gov/27031506/)]
37. Guiot HM, Franqui-Rivera H. Predicting performance on the United States Medical Licensing Examination Step 1 and Step 2 Clinical Knowledge using results from previous examinations. *AMEP* 2018 Dec;Volume 9:943-949. [doi: [10.2147/amep.s180786](https://doi.org/10.2147/amep.s180786)]
38. USMLE Score Interpretation Guidelines. United States Medical Licensing Examination. URL: [http://www.usmle.org/pdfs/transcripts/USMLE\\_Step\\_Examination\\_Score\\_Interpretation\\_Guidelines.pdf](http://www.usmle.org/pdfs/transcripts/USMLE_Step_Examination_Score_Interpretation_Guidelines.pdf) [accessed 2022-08-18]
39. Osman A, Barrios FX, Kopper BA, Hauptmann W, Jones J, O'Neill E. Factor structure, reliability, and validity of the Pain Catastrophizing Scale. *J Behav Med* 1997 Dec;20(6):589-605. [doi: [10.1023/a:1025570508954](https://doi.org/10.1023/a:1025570508954)] [Medline: [9429990](https://pubmed.ncbi.nlm.nih.gov/9429990/)]
40. Shi Y, Gugiu PC, Crowe RP, Way DP. A Rasch Analysis Validation of the Maslach Burnout Inventory-Student Survey with Preclinical Medical Students. *Teach Learn Med* 2019 Dec 21;31(2):154-169. [doi: [10.1080/10401334.2018.1523010](https://doi.org/10.1080/10401334.2018.1523010)] [Medline: [30577705](https://pubmed.ncbi.nlm.nih.gov/30577705/)]
41. Cohen S, Williamson GM. Perceived Stress in a Probability Sample of the United States. In: *The Social Psychology of Health*. Newbury Park, CA: Sage Publications; 1988:31-67.
42. 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](https://doi.org/10.2307/2136404)]
43. Bollen KA, Hoyle RH. Perceived Cohesion: A Conceptual and Empirical Examination. *Social Forces* 1990 Dec;69(2):479. [doi: [10.2307/2579670](https://doi.org/10.2307/2579670)]
44. West CP, Shanafelt TD, Kolars JC. Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. *JAMA* 2011 Sep 07;306(9):952-960. [doi: [10.1001/jama.2011.1247](https://doi.org/10.1001/jama.2011.1247)] [Medline: [21900135](https://pubmed.ncbi.nlm.nih.gov/21900135/)]
45. Levin R, Cohen J, White L. Gold Humanism Honor Society Membership Indicator to Be Added to ERAS Applications Beginning With 2016 Application Cycle. *J Grad Med Educ* 2015 Mar;7(1):136 [FREE Full text] [doi: [10.4300/JGME-D-14-00723.1](https://doi.org/10.4300/JGME-D-14-00723.1)] [Medline: [26217447](https://pubmed.ncbi.nlm.nih.gov/26217447/)]
46. Buck E, Holden M, Szauter K. A Methodological Review of the Assessment of Humanism in Medical Students. *Academic Medicine* 2015;90:S14-S23. [doi: [10.1097/acm.0000000000000910](https://doi.org/10.1097/acm.0000000000000910)]

## Abbreviations

**GHHS:** Gold Humanism Honors Society  
**HIMI:** Humanism in Medicine Initiative  
**MBI:** Maslach Burnout Inventory  
**OSCE:** Objective Structured Clinical Examination  
**OSUCOM:** Ohio State University College of Medicine  
**PCS:** Perceived Cohesion Scale  
**PSS:** Perceived Stress Scale  
**QOL:** quality of life  
**USMLE:** United States Medical Licensing Examination

*Edited by A Mavragani; submitted 12.02.22; peer-reviewed by KH Tan, L Puljak; comments to author 25.04.22; revised version received 11.08.22; accepted 15.08.22; published 16.09.22.*

*Please cite as:*

*Auckley ED, Barbee J, Verbeck N, McCambridge T, Stone L, Garvin J*

*Extracurricular Humanism in Medicine Initiative and Medical Student Wellness: Retrospective Study*

*JMIR Form Res 2022;6(9):e37252*

*URL: <https://formative.jmir.org/2022/9/e37252>*

*doi: [10.2196/37252](https://doi.org/10.2196/37252)*

*PMID: [36112401](https://pubmed.ncbi.nlm.nih.gov/36112401/)*

©Elizabeth Diane Auckley, Jeff Barbee, Nicole Verbeck, Tracie McCambridge, Linda Stone, Jennifer Garvin. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 16.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Viewpoint

# Collaborative Challenges of Multi-Cohort Projects in Pharmacogenetics—Why Time Is Essential for Meaningful Collaborations

Filippo Franchini<sup>1</sup>, MSc; Katharina Kusejko<sup>2</sup>, DPhil; Catia Marzolini<sup>3</sup>, MD; Christoph Tellenbach<sup>4</sup>, DPhil; Simona Rossi<sup>5</sup>, MSc; Susanne Stampf<sup>5</sup>, DPhil; Michael Koller<sup>5</sup>, MD; Jivko Stoyanov<sup>6,7</sup>, PhD; Burkhard Möller<sup>8</sup>, MD; Alexander Benedikt Leichtle<sup>1,9</sup>, MD

<sup>1</sup>University Institute of Clinical Chemistry, Inselspital - Bern University Hospital, University of Bern, Bern, Switzerland

<sup>2</sup>Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zürich, Switzerland

<sup>3</sup>Division of Infectious Diseases and Hospital Epidemiology, University Hospital Basel, University of Basel, Basel, Switzerland

<sup>4</sup>Swiss Clinical Quality Management in Rheumatic Diseases, Zürich, Switzerland

<sup>5</sup>Swiss Transplant Cohort Study, Clinic for Transplantation Immunology and Nephrology, University Hospital Basel, Basel, Switzerland

<sup>6</sup>Swiss Paraplegic Research, Nottwil, Switzerland

<sup>7</sup>Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

<sup>8</sup>University Clinics of Rheumatology and Immunology, Inselspital - Bern University Hospital and University of Bern, Bern, Switzerland

<sup>9</sup>Center for Artificial Intelligence in Medicine, University of Bern, Bern, Switzerland

**Corresponding Author:**

Alexander Benedikt Leichtle, MD  
University Institute of Clinical Chemistry  
Inselspital - Bern University Hospital  
University of Bern  
Freiburgstrasse 10  
Bern, 3010  
Switzerland  
Phone: 41 31 6328330  
Email: [Alexander.Leichtle@insel.ch](mailto:Alexander.Leichtle@insel.ch)

**Abstract**

Multi-cohort projects in medicine provide an opportunity to investigate scientific questions beyond the boundaries of a single institution and endeavor to increase the sample size for obtaining more reliable results. However, the complications of these kinds of collaborations arise during management, with many administrative hurdles. Hands-on approaches and lessons learned from previous collaborations provide solutions for optimized collaboration models. Here, we use our experience in running PGX-link, a Swiss multi-cohort project, to show the strategy we used to tackle different challenges from project setup to obtaining the relevant permits, including ethics approval. We set PGX-link in an international context because our struggles were similar to those encountered during the SYNCHROS (SYnergies for Cohorts in Health: integrating the ROle of all Stakeholders) project. We provide ad hoc solutions for cohorts, general project management strategies, and suggestions for unified protocols between cohorts that would ease current management hurdles. Project managers are not necessarily familiar with medical projects, and even if they are, they are not aware of the intricacies behind decision-making and consequently, of the time needed to set up multi-cohort collaborations. This paper is meant to be a brief overview of what we experienced with our multi-cohort project and provides the necessary practices for future managers.

(*JMIR Form Res* 2022;6(9):e36759) doi:[10.2196/36759](https://doi.org/10.2196/36759)

**KEYWORDS**

personalized medicine; guidelines; ethical, legal, and social implications; study; ethics; multicentric

## Background

Multi-cohort studies are increasingly important in medicine as they provide the opportunity to join data and join efforts to increase the sample size and investigate questions beyond the scope of a single institution. Multi-cohort projects not only encourage cross-boundary collaborations, but they also boost synergies between cohorts by providing staff and funding. There are multi-cohort projects dealing with complementary cohorts around a similar subject and projects dealing with generalist cohorts, but none of them are representative of the general population.

At the international level, there are many initiatives that aim to facilitate collaborations between cohorts. The SYNCHROS (SYnergies for Cohorts in Health: integrating the ROLE of all Stakeholders) project, a 3-year project funded by the Horizon 2020 Program that started in 2019, successfully mapped 1000 multi-cohort projects in 11 countries with the intent to harmonize the coordination and data interoperability of multi-cohort projects in Europe and worldwide [1,2]. SYNCHROS has a focus on identifying the practical, methodological, ethical, and legal challenges that cohorts are facing and on ideas how to facilitate the development toward tapping in data, which can be valuable for personalized medicine. We also encountered similar challenges while setting up our multicentric and multi-cohort project.

In this paper, we focus on different cohorts dealing with intrinsically different patients who are complementary but form a heterogeneous multi-cohort environment. The project tries to answer a particular scientific question by using patients selected from that heterogeneous multi-cohort environment. Finding a common ground across multiple cohorts and a consensus to a particular multi-cohort scientific question can be difficult. Furthermore, different cohorts have specific purposes, focus areas, and policies as well as their own established methods of managing, collecting, and sharing data.

Here, we use the Swiss project, PGX-link, as an exemplary case study to show the complexity of setting up a 2-year multi-cohort project—in our case, in the field of pharmacogenetics. This project is a feasibility study, focused on building an infrastructure that connects clinical and pharmacogenomics data among 3 cohorts: the Swiss HIV Cohort Study (SHCS) [3], the Swiss Transplant Cohort Study (STCS) [4,5], and the Swiss Clinical Quality Management (SCQM) in rheumatic disease [6]. The 3 cohorts were chosen because they previously worked together. In the COVERALL (Corona VaccinE tRiAL pLatform) study, patients from STCS and SHCS were vaccinated for the purpose of a research project and sample shipment was coordinated between the cohorts [7]. The IDEAL (IDentifier LinkAge between patient-centered research and heaLthcare providers) project, including STCS, SHCS, and SCQM, focuses on automatic ways of identifying patients with their cohort IDs within the hospital system [8]; this is crucial for improving the

matching, and as a result, makes data extraction from the hospital systems more feasible. New applications for interoperable “National Data Streams” within the Swiss Personalized Health Network (SPHN) and Swiss National Science Foundation (SNSF) funding framework also aim to use a multi-cohort approach. Therefore, PGX-link is not the only effort toward more interoperability and improvement of data flow, and it benefited a lot from the fact that the 3 cohorts were not completely “strangers.” However, in the past, the 3 cohorts never joined forces for a study like PGX-link (a genetic study consisting of the development of an infrastructure to answer a scientific question) and, as a consequence, the process of obtaining the relevant approvals was never explored at a multi-cohort level.

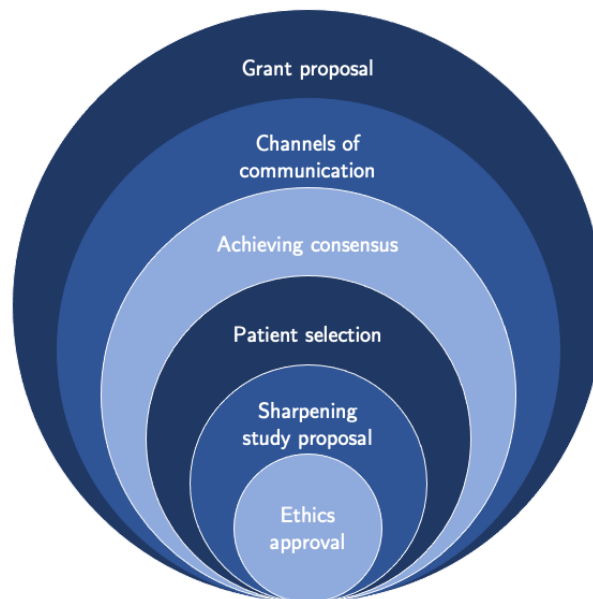
The project infrastructure was funded by the SNSF within the context of the BioLink program fostering collaborations between cohorts and their biobanks. The laboratory work and genetic analyses were funded by the Bern Centre for Precision Medicine (BCPM). Because of their large size, multi-centric projects are often funded by different bodies. PGX-link is a good example of a multi-cohort project where the access to the 2 different fundings was bounded to different prerequisites that were not met at the same time. This added an additional layer of complexity.

We show the evolution of the study protocol from the initial SNSF grant proposal to the final submission to the cohorts’ scientific boards and ethics committees to receive data and samples. We highlight the obstacles that we encountered in setting up and managing PGX-link as a multi-cohort collaboration. We explain in detail the changes in the proposal necessary for obtaining the project approval by the ethics committee, and we provide suggestions and potential solutions that can improve and support such processes for future multi-cohort projects and likewise serve as a guidance for funding agencies. Although some of the examples that we will provide here are “Swiss-centric,” we believe that the underlying messages that such examples convey as well as the intracohort and intercohort dynamics that we describe can be useful also at an international level.

## The Fundamental Steps of the Project

For PGX-link, we used an approach to answer a specific scientific question relying on a multi-cohort framework in a unique and ad hoc way. This strategy was preferred to using an infrastructure approach that would first set up an infrastructure and later on define the scientific questions that can be answered using the setup. Note that although from the cohort perspective, the scientific question approach adopts an ad hoc strategy, the question we want to answer acts as a common denominator across the cohorts joining the cohorts together. Thus, in the early stage of the PGX-link project, we went through the following 6 steps (see [Figure 1](#)):

**Figure 1.** Consecutive spheres of collaborative endeavor of the PGX-link project from the grant proposal write-up through reaching intercohort consensus and the final ethics approval.



1. Write-up of the grant proposal for the first funding body (SNSF) to fund the project infrastructure.
2. Establishing channels of communication with the cohorts.
3. Achieving consensus and feasibility discussion on a multi-cohort scientific question.
4. Patient selection within each cohort.
5. Sharpening of study proposal for the scientific boards.
6. Finalization of the study protocol and collection of the relevant documents to receive the approval from the ethics committee. The ethics committee approval is needed to retrieve funding from the second funding body (BCPM).

Although these steps might be slightly different or be in a different order in other countries, projects in medicine using software quality assurance will eventually require a cocktail of these steps to enable the retrieval of data and samples. Consensus on the scientific question is achieved after establishing the channel of communication with the cohorts. The scientific question is then used to select the patients, obtain approval from scientific boards and ethics committees, and access the BCPM funding.

At the end of each step, we critically reviewed our approach and gave suggestions to facilitate knowledge translation and successful collaboration. These steps formed a preparation phase that we used to set up all the “fundamental bricks” necessary to start the project. These bricks consist of compulsory paperwork to obtain the project approval from the cohorts and the ethics committees. As this long process includes sensitive data, it involves many people at different decision levels. Therefore, the preparation phase of a multi-cohort study takes a substantial amount of time (in our case, 1 year). This is a detail that is not obvious for funding bodies, which, in turn, are dependent on external assessment focusing on the scientific return on investment.

Cohort members and the ethics committee were extremely helpful and provided essential support to obtain the ethics approval necessary to start the project. Nevertheless, as this was a first-of-its-kind study in Switzerland, the bureaucratic process from step 1 to step 6 was not yet clearly perceivable. We thus decided to describe our journey and provide guidance for future multi-cohort studies. Note that, although the process described here might be transferable to other multi-cohort projects at national and international levels, some of the strategies that we used might not apply to other cohorts and regulatory circumstances.

### *Step I: Writing the Grant Proposal*

Before diving into the project evolution, it is important to know that each cohort has its own data/sample governance structure. The scientific board is responsible for evaluating the project proposal, especially the scientific question that the project aims to investigate. The foundation board is responsible for expressing the support to the project, and it includes legal representatives having the authority to sign documents on behalf of the cohort.

An ethics approval was necessary to receive the funding from BCPM and officially start the project (as such, funding was necessary to perform the genetic analysis). The ethics approval was conditional on approval from the scientific boards. In our case, p-v1 (proposal version 1, the first version of the proposal submitted to the SNSF) was approved by the foundation boards supporting data and sample sharing within our multi-cohort initiative, but it was not approved by the scientific boards. The SNSF BioLink grant specifically funded the infrastructure of the project. Still, a scientific question was required for successful submission, but it was not a key aspect of the grant. This scientific question was added to p-v1 close to its submission on

the data available at this time and thus was not reviewed by the scientific boards. Hence, the evaluation of the scientific question by the scientific boards was independent of the support of the foundation boards.

The scientific boards were not involved because of organizational incompatibilities: it was not possible to receive feedback from the cohorts' collaborators while writing the grant because deadlines were extremely short and during typical summer holiday times. Without in-depth knowledge of the cohort data and proper discussion between the cohorts themselves, the scientific question could only be superficially outlined. This explains why after receiving the SNSF grant, the scientific question in p-v1 was not entirely clear to the scientific boards. The scientific boards would only agree on p-v1 if the scientific question was clearly stated and was precise and feasible with enough patients available to guarantee the statistical power sufficient for genetic studies. Having enough time between grant call and deadlines to involve collaborators early on would have allowed the project management to be more efficient. Following the feedback of the scientific boards and the cohort collaborators, the proposal had to be thoroughly revised. We thus used a substantial amount of the project time and funding just for the proper preparation without even starting the actual project. To reach a consensus on a multi-cohort scientific question (ie, a feasible question that is of interest to all 3 cohorts), we set up the intercohort collaboration network. This led to discussions with cohort collaborators and scientific boards that helped to understand that the scientific question proposed in p-v1 was not sufficiently fitting the available data and too complicated for a genetics study within the given timeframe.

We experienced that not having the scientific question clearly defined from the beginning is detrimental to the project's budget and timeline. Thus, for a multi-cohort project to be efficient, it is important to involve scientific board members and cohort collaborators early during the writing phase of the grant proposal—at least 1-2 people from each cohort who are experts in the field, who know “their” cohort's data, and who can contribute at an early stage. Asking the funding bodies to have an early announced application process with an adequate duration of the presubmission phase is crucial for the proper allocation of people and resources. It is important to have sufficient time between the grant call and deadline to allow for appropriate communication between partners and for the scientific boards to adapt to the process. In this way, the proposal can benefit from feedback from the scientific boards.

## ***Step II: Setting Up Channels of Communication***

In this step, we opened channels of communications between the cohorts, establishing the intercohort collaboration network, as well as with laboratory services, information technology services, and legal services to address all concerns from the cohorts and the ethics committee. For PGX-link, we had expert collaborators with different backgrounds on board: physicians, pharmacologists, and representatives of the cohorts involved in data management. This provided the project manager with strong

scientific and practical support during discussions. The collaborators were approximately 10 people who were mostly active during the elaboration of the reviewed version of p-v1, that is, proposal v2.0 or p-v2. The implementation of the scientific boards' feedback and the inclusion of detailed information collected from information technology services (data security and storage), laboratory services (costs and procedures to analyze the samples), and legal services (data and material transfer agreements) led to the official version of p-v2 ready to be submitted to the ethics committee.

An official review and a thorough discussion of the project are usually done during scientific board meetings that are conducted approximately 4 times a year. We would like to make the reader aware that cohort collaborators in the project are not necessarily part of the cohort scientific board committee. Involving only non-scientific board members would result in a proposal that does not address the concerns of the scientific board. This could lead to multiple review rounds, thereby increasing preparation time. In our case, we went through a minimum of 2 rounds in each cohort. Thus, to be time-efficient, we recommend that the project manager includes scientific board members at an early stage in the project so that (1) early concerns of the scientific board can be addressed and (2) project updates can be communicated beforehand to the other scientific board members. Multi-cohort projects depend on multiple scientific boards (one for each cohort) that within a year have a few meetings at different time points. Instead of meeting the scientific boards separately and dramatically increasing the preparation time, we encourage involving members of other cohorts' scientific boards to the first available scientific board meeting. In this way, it is possible to obtain a more holistic view of the project and address multi-cohort concerns in 1 meeting. If not already standard practice, we also encourage the project management to participate in the scientific board meetings when the scientific board is discussing your project. This is the opportunity to present the project in person and answer upcoming questions. Besides these suggestions, we recommend standard best practices in project management such as (1) avoiding planning short notice meetings and using a widely accepted meeting planning software to plan meetings, (2) keeping an agenda of all the meetings, and (3) communicating the meeting agenda in advance and sharing the minutes shortly afterwards. Not following such points could result in collaborators who do not know who is attending the meeting, who do not find time to participate, who do not get an update where the project stands, and what the next steps are or what contribution is expected from them.

The sharpening of the scientific question was probably the most difficult task in the preparation phase of PGX-link, requiring all the cohorts to engage in discussion and find a consensus. Multi-cohort meetings created an occasion for joint discussions, making collaborators aware of what is collected in the other cohorts and adjust the question accordingly.



### ***Step III: Sharpening the Scientific Question***

The number of prospective patients is necessary to demonstrate the feasibility of the study in terms of participant numbers needed for the statistical power and budget. Having the correct number of patients implies that the scientific question is clear and precise so that it is possible to define the inclusion and exclusion criteria for the patient selection. If there are not enough patients, the criteria need to be iteratively rediscussed and possibly adapted until a sufficient number is met (as the number of cohort participants cannot easily be expanded).

In the original grant proposal p-v1 and previously in the funding offer from the SNSF, the term “infrastructure” was kept relatively open, probably because in earlier stages, there was not enough knowledge about the general feasibility of such a study. Thus, as a feasibility study, instead of building an exhaustive infrastructure, we decided to devise a small cohort metadata set containing only the data necessary to answer the scientific question. The cohort metadata set can be explored using some raw code that can be viewed as core infrastructure (ie, a raw code reviewed by cohort collaborators that can be used to perform, eg, patient selection within a specific cohort). This can be scaled to include more data and allow a more user-friendly interaction. Such evolution was embraced by the cohorts as a more realistic approach than those initially proposed in p-v1.

With PGX-link, we focused on chronic kidney disease in patients mainly treated with antiviral or immunomodulatory medications that are known to bear a risk of renal impairment and investigated the genetic variants that are predictive of a decline in renal function. Some crucial technical and epidemiological assumptions/prerequisites had to be met: (1) common medication, (2) availability of creatinine measurements results and dates, (3) availability of biosamples, and (4) semantically interoperable data (or metadata). Points 1-3 revealed some real-world problems as patients within the different cohorts had different diseases. However, such scientific questioning presented important advantages. We used estimated glomerular filtration rate as the criterion to select patients for this study, as this value can be simply calculated using sex, age, and serum creatinine levels of the patients [9,10]. Creatinine levels are widely measured and accessible. Thus, we only needed a few basic data items that are routinely collected. Moreover, the existence of previous studies [11,12] focusing on renal impairment in patients within the same cohorts allowed us to (1) already have genetic data available and thus increase the number of patients, (2) have peer-reviewed methodology for patient selection by using estimated glomerular filtration rate values, and (3) have support from previous collaborators involved in patient selection and genetic analysis.

Thus, we recommend sticking with common best practices in project management. It is crucial that the right terminology and clear communication are used when we describe a project to cohorts. We also encourage to keep the study as simple as possible. Unexpected problems are always around the corner. This is important for any study and even more important for

multi-cohort studies. Whenever possible, available methodologies or data should be used, as it will save time and resources. An additional hurdle is the fact that templates for letter of intent and full proposal could differ between the cohorts. If this is the case, it would be more efficient to have a standardized template for multi-cohort studies in order not to submit the same content multiple times in different formats.

### ***Step IV: Defining the Number of Participants***

The number of participants is a crucial factor for the statistical power of a study and to obtain the approval from the scientific boards and the ethics committee. A clear scientific question and clear criteria are necessary for selecting the patients for a study. However, without the approval from the ethics committee (dependent on scientific board approval), it is not possible to obtain access to patients' data. Therefore, how can we estimate the number of suitable patients if we do not have access to the data at the time of writing the proposal?

The approach we took was using remote patient selection in which the cohorts provide the structure of their database to the project manager who, in turn, devises a piece of software (in our case, an R-package on GitHub called PGX-link tools [13]) that the cohorts can use and run on their local secured infrastructure to provide a rough estimate of the number of potential patients. Although this step worked well, it required tight cooperation between the project manager and cohorts, since not having direct access to the data makes the development of the package more difficult. This procedure allows only broad estimates of the number of participants because the package might contain bugs that can be fixed only once the developer has access to the data, and additional exploratory analysis can reveal that some selected patients are nevertheless not eligible for the study. Still, in most cases, remote patient selection is precise enough to obtain the required approvals from the scientific boards and the ethics committee, with the caveat that numbers after final patient selection might differ. We suggest that the developer should access the data before submitting the proposal, as it increases precision in patient selection. Currently, the only way of doing that is by being physically present in the cohorts' infrastructure and supported by a team member. The development of a tool to interact with cohorts' data requires some time to devise. It is not only dependent on the developer but also on the cohorts' collaborators, especially data managers and clinicians, which proved to be extremely supportive in our case. In the best case scenario, such a developmental stage requires probably 1-2 weeks and in the worst-case scenario, even up to 1-2 months. In our case, the cohort dealing with patients who underwent transplantation (STCS) presented a particular example since the time from transplantation to case definition (ie, follow-up period) added an additional layer of complexity. Another recently released option to perform remote patient selection could be a so-called cohort explorer [14], that is, a secured web-based infrastructure that can be used to explore data within a specific cohort. However, depending on the granularity of the queries, this might not always be sufficient.

Usually, cohort patients treated in Swiss hospitals have their data stored in the hospitals' databases (it is even required by law that the hospitals keep the patient data records). Such data can include medication and routine measurements such as creatinine levels. In our case, the creatinine level data of 1 cohort were kept in 2 Swiss hospitals. To retrieve those creatinine values, additional paperwork was required to obtain the relevant permits. Be aware that there could be problems of underreporting: in our specific case, medication information retrieved from 1 hospital was not as detailed as that retrieved within the cohorts. It is thus important to double check with the cohorts any data that are retrieved externally, as the cohorts generally have more cautiously curated databases.

Regarding the analysis of biosamples, cohorts have their own policies. If a genetic database is already in place, it is very likely that they will prevent patchwork analyses from different projects but rather analyze many samples that will include most but not all samples needed for the project. Thus, it is important to have enough funding to participate in large genetic screenings that include samples from multiple projects, with the aim of getting your samples analyzed. Cohorts will always be happy to generate more genetic data for their patients to have a broader data foundation that can be potentially reused by other projects. However, from a project management point-of-view, the benefit of obtaining additional data from genetic analysis must outweigh the costs. If the cohorts already have genetic data available, it is good practice to use those data first, and then if strictly required, proceed with additional screening.

Thus, in this section, we recommend, whenever possible, to avoid remote patient selection and visit the cohorts to perform patient selection with actual patient data without working with mock data sets. We also encourage to use genetic data that are already available before proceeding with additional genetic analyses as they are expensive and are usually prone to substantial time delay especially if the cohorts are putting together multiple projects to screen a large number of patients.

### ***Steps V and VI: Obtaining Scientific Board and Ethics Committee Approvals***

Two perspectives are intertwined in the progress of complex collaborations—reaching sound and consensual strategic decisions and timely decisions. The latter does not have to always wait for the former—the project timeline can benefit from a proactive anticipatory approach in which certain milestones are de facto prepared and ready to be used once a decision is reached on a strategic level. Many steps are strictly dependent on the previous ones, sometimes with insolvable circularities. Although these dependencies are in place to ensure security when dealing with patients' sensitive data, scientific board and ethics committee approvals can be overlapping when being addressed jointly. The draft of p-v2 was already included in the ethics committee submission to gain time and to obtain the conditional ethics committee approval without waiting for the scientific board meetings. Thus, we obtained the conditional ethics committee approval before the proposal was officially accepted by all scientific boards (1 cohort officially accepted before submission to the ethics committee). With the conditional

ethics committee approval, we were able to obtain the hospital approval to start the data search with the help of the data science services. Additionally, the conditional ethics committee approval probably increased the trust in the project by the cohorts.

Thus, we suggest not to wait for all the scientific boards to formally approve the final project plan. The application to the ethics committee should be started as soon as a viable version of the proposal has been approved by the collaborators and possibly contains some feedback from the cohort scientific boards. Ethics committee clearance is a formal and official process that usually takes some time. The approval will ease (but not speed up) the processes within the cohorts. Amendments to the protocols and ethics submissions might be necessary but are simpler *ex posteriori*.

### ***What to “Take Home”***

In this paper, we illustrate how complex the preparation phase of a multi-cohort project in pharmacogenomics can be. We encountered multiple challenges along the way, and we believe they will likely affect other researchers conducting similar projects at national and international levels in the future as they have been already encountered in the past [1,2]. We expect that in the future, the number of funding opportunities for such projects will rise (compared to monocohort or monocentric projects) and these projects will provide opportunities to highlight current heterogeneities between different cohorts but likewise, unlock their great collaborative potentials. The SYNCHROS project was a €2 million research program that mapped 1000 multi-cohort projects funded by European, American, and Canadian institutions. Although the project is at its end, it is likely that such type of grants will continue in the future.

It is clear to us that heterogeneities between cohorts will always exist and remain because of the different nature of the cohort-specific diseases. However, there are some minor adaptations that can facilitate the future fluidity of multi-cohort studies. For example, each cohort has its own specific material and data transfer agreement, own proposal template, and own way of approaching data sharing and security. These challenges were previously mentioned at the international level, and the harmonization of such documents would ease collaboration at the national and international levels [15,16]. If all cohorts will share a common proposal template, material and data transfer agreement, and the system of sharing data, the processes will be simpler and seamless between the cohorts. Of course, this will require efforts not only from the cohorts but also from legal services involved in such types of legal contracts. We propose a unified template for the letter of intent and for a full proposal for all the national cohorts, while legal services could provide the material and data transfer agreement based on the centers involved in the project [17].

We noticed that despite national initiatives to improve synergies such as the Swiss Biobanking Platform (SBP) and SPHN, at the time of writing the proposal, the cohorts were still in the process of obtaining the newly released biobank labels of SBP or having their database adapted, as recently suggested by SPHN. Most of the cohorts are much longer established than

SBP and SPHN and, as a consequence, such restructuring processes can interfere with the cohorts' timelines, and more importantly, be very expensive. The current situation does not yet allow cohorts to invest time and resources in changing their system in favor of potential further collaborations, and money should be specifically allocated for this task. For example, the application process to obtain the quality label from SBP is detailed and requires help from laboratory members who are not paid by the cohort and most of the time help "in kind." However, the immediate benefit of obtaining that label is that the cohort biobank has been checked and internal procedures such as protocols and type of laboratory equipment have been reviewed. Researchers that use biosamples from labeled biobanks can be sure that the samples they are using follow SBP standards, and this makes their methodology solid for potential publications. The management of biobanks at a multi-cohort level is essential to ease the process of patient selection or simply as an exploration tool for project management. This has been previously emphasized at the national and international levels to improve method harmonization and integration [16]. The extent of overlap and standardization between the cohorts will prospectively rise with specific funding for multi-cohort projects. Hence, funding agencies should incorporate the intricacies of this type of collaboration in their funding schemes.

Ideally, the cohorts increase synergies by starting regular multi-cohort communications, following the new standards of SBP and SPHN, and agreeing on common document templates such as letters of intent, project proposals, and data transfer agreements. Common variables (or metadata definitions) across the cohorts ensure a minimum viable meta-cohort data set that can be used in multi-cohort projects, especially in short-term feasibility studies. Such a data set can include routinely collected data that are common in all cohorts, such as gender, age, weight, height, medication (eg, via Anatomical Therapeutic Chemical codes), blood pressure, and standardized laboratory assessments.

The feasibility of having a minimum viable data set needs discussion, as scientific questions are usually specific, and only with basic information, one cannot spark interesting projects. If these small steps are in place before a multi-cohort project starts, from the cohorts' perspective, one would avoid patchwork from multiple projects requesting different data just to answer a particular question. In contrast, from the project management perspective, filling the same type of documents multiple times and performing multiple submissions of similar documents could be avoided. To make this happen, funding bodies must foster the use of common standards for all cohorts and funding must be available nationwide to ease the transition from a monocentric standpoint to a multi-cohort perspective. Furthermore, currently, little importance is given to the administrative, regulatory, and scientific setup of a multi-cohort that easily could take up to 1 year. The SYNCHROS project concludes that retrieving data, analyzing them, and publishing findings are not the only challenges in a multi-cohort project. A lot of funding and efforts must be devoted to address practical, methodological, ethical, and legal challenges that usually arise before the actual project starts, as there are existing infrastructures and governance variants for each involved cohort. Here, we suggest that it would be helpful if funding bodies could divide the grant that can be used to set up the project before it officially starts (ie, getting all the necessary approvals and clearance to sensitive data) from a grant to actually conduct the project. It is thus important that the scientific value of a project is clearly outlined and peer reviewed before proceeding with its execution. In our view, funding only the infrastructure is not enough, as an infrastructure must have a clear purpose and answer specific questions. We suggest that funding bodies should request approval from the scientific boards before releasing the grant. This will ensure that substantial additional preparatory work has been done by the grantees, as communication with cohorts is the key to get the scientific boards on board as we outlined in this paper.

---

## Acknowledgments

The PGX-link project was funded by the Swiss National Science Foundation (project ID 2021-01104) and by the Bern Centre for Precision Medicine (PGX-link PGM). The authors would like to thank Dr Aurélie Pahud de Mortanges for her critical and constructive review of the manuscript, thereby rendering it easy to read. The authors are also grateful to Dr Almut Scherer and Dr Guido Stirnimann for their excellent comments that greatly improved the manuscript.

---

## Conflicts of Interest

None declared.

---

## References

1. Rodríguez-Laso Á, Rico-Urbe LA, Kubiak C, Haro JM, Rodríguez-Mañas L, Ayuso JL. A Map of the Initiatives That Harmonize Patient Cohorts Across the World. *Front Public Health* 2021;9:666844 [FREE Full text] [doi: [10.3389/fpubh.2021.666844](https://doi.org/10.3389/fpubh.2021.666844)] [Medline: [34336764](https://pubmed.ncbi.nlm.nih.gov/34336764/)]
2. Main European and international efforts to harmonise and integrate cohort data. SYNCHROS Repository. URL: <https://repository.synchros.eu> [accessed 2022-08-29]
3. Scherrer AU, Traytel A, Braun DL, Calmy A, Battegay M, Cavassini M, Swiss HIV Cohort Study (SHCS). Cohort Profile Update: The Swiss HIV Cohort Study (SHCS). *Int J Epidemiol* 2022 Feb 18;51(1):33-34j. [doi: [10.1093/ije/dyab141](https://doi.org/10.1093/ije/dyab141)] [Medline: [34363666](https://pubmed.ncbi.nlm.nih.gov/34363666/)]

4. Koller MT, van Delden C, Müller NJ, Baumann P, Lovis C, Marti H, et al. Design and methodology of the Swiss Transplant Cohort Study (STCS): a comprehensive prospective nationwide long-term follow-up cohort. *Eur J Epidemiol* 2013 Apr;28(4):347-355 [FREE Full text] [doi: [10.1007/s10654-012-9754-y](https://doi.org/10.1007/s10654-012-9754-y)] [Medline: [23546766](https://pubmed.ncbi.nlm.nih.gov/23546766/)]
5. Stampf S, Mueller NJ, van Delden C, Pascual M, Manuel O, Banz V, Swiss Transplant Cohort Study, members of the Swiss Transplant Cohort Study. Cohort profile: The Swiss Transplant Cohort Study (STCS): A nationwide longitudinal cohort study of all solid organ recipients in Switzerland. *BMJ Open* 2021 Dec 15;11(12):e051176 [FREE Full text] [doi: [10.1136/bmjopen-2021-051176](https://doi.org/10.1136/bmjopen-2021-051176)] [Medline: [34911712](https://pubmed.ncbi.nlm.nih.gov/34911712/)]
6. Langenegger T, Fransen J, Forster A, Seitz M, Michel BA, Swiss Clinical Quality Management in Rheumatoid Arthritis. [Clinical quality management in rheumatoid arthritis]. *Z Rheumatol* 2001 Oct;60(5):333-341. [doi: [10.1007/s003930170033](https://doi.org/10.1007/s003930170033)] [Medline: [11759233](https://pubmed.ncbi.nlm.nih.gov/11759233/)]
7. Speich B, Chammartin F, Smith D, Stoeckle MP, Amico P, Eichenberger AL, study groups from the Swiss HIV Cohort Study, the Swiss Transplant Cohort Study. A trial platform to assess approved SARS-CoV-2 vaccines in immunocompromised patients: first sub-protocol for a pilot trial comparing the mRNA vaccines Comirnaty® and COVID-19 mRNA Vaccine Moderna®. *Trials* 2021 Oct 21;22(1):724 [FREE Full text] [doi: [10.1186/s13063-021-05664-0](https://doi.org/10.1186/s13063-021-05664-0)] [Medline: [34674742](https://pubmed.ncbi.nlm.nih.gov/34674742/)]
8. Koller M, Scherrer A. About the IDEAL project. IDEAL. URL: <https://ideal-project.ch/about/> [accessed 2022-08-29]
9. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, et al. New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race. *N Engl J Med* 2021 Nov 04;385(19):1737-1749. [doi: [10.1056/nejmoa2102953](https://doi.org/10.1056/nejmoa2102953)]
10. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009 May 05;150(9):604-612 [FREE Full text] [doi: [10.7326/0003-4819-150-9-200905050-00006](https://doi.org/10.7326/0003-4819-150-9-200905050-00006)] [Medline: [19414839](https://pubmed.ncbi.nlm.nih.gov/19414839/)]
11. Dietrich LG, Barceló C, Thorball CW, Ryom L, Burkhalter F, Hasse B, et al. Contribution of Genetic Background and Data Collection on Adverse Events of Anti-human Immunodeficiency Virus (HIV) Drugs (D:A:D) Clinical Risk Score to Chronic Kidney Disease in Swiss HIV-infected Persons With Normal Baseline Estimated Glomerular Filtration Rate. *Clin Infect Dis* 2020 Feb 14;70(5):890-897. [doi: [10.1093/cid/ciz280](https://doi.org/10.1093/cid/ciz280)] [Medline: [30953057](https://pubmed.ncbi.nlm.nih.gov/30953057/)]
12. Dietrich L, Thorball C, Ryom L, Burkhalter F, Hasse B, Thurnheer M, Swiss HIV Cohort Study. Rapid Progression of Kidney Dysfunction in People Living With HIV: Use of Polygenic and Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) Risk Scores. *J Infect Dis* 2021 Jun 15;223(12):2145-2153. [doi: [10.1093/infdis/jiaa695](https://doi.org/10.1093/infdis/jiaa695)] [Medline: [33151293](https://pubmed.ncbi.nlm.nih.gov/33151293/)]
13. Franchini F. PGX-link tools. GitHub. URL: <https://github.com/FilippoFranchini/pgxTools> [accessed 2022-08-29]
14. The SPHN Federated Query System. URL: <https://sphn.ch/2021/11/16/sphn-fqs-launch/> [accessed 2022-08-29]
15. Bickenbach J, Sialm A. Strategy brief on ethical and legal issues in the optimization of cohort data in Europe. SYNCHROS. URL: <https://synchros.eu/wp-content/uploads/Full-doc-SB-Ethics.pdf> [accessed 2022-08-29]
16. Sanchez-Niubo A, Sialm A. Strategy brief on harmonization and integration methods. SYNCHROS. URL: <https://synchros.eu/wp-content/uploads/Full-doc-SB-Methods-1.0.pdf> [accessed 2022-08-29]
17. Reutimann H. The Swiss Biotech Association (SBA) and Unitecra, the Joint Tech Transfer Organization of the Universities of Basel, Bern and Zurich - Two Organizations Initiated and Fostered by Oreste Ghisalba. *Chimia (Aarau)* 2020 May 27;74(5):360-362. [doi: [10.2533/chimia.2020.360](https://doi.org/10.2533/chimia.2020.360)] [Medline: [32482211](https://pubmed.ncbi.nlm.nih.gov/32482211/)]

## Abbreviations

**BCPM:** Bern Centre for Precision Medicine

**SBP:** Swiss Biobanking Platform

**SCQM:** Swiss Clinical Quality Management

**SHCS:** Swiss HIV Cohort Study

**SNSF:** Swiss National Science Foundation

**SPHN:** Swiss Personalized Health Network

**STCS:** Swiss Transplant Cohort Study

**SYNCHROS:** SYnergies for Cohorts in Health: integrating the ROle of all Stakeholders

*Edited by A Mavragani; submitted 24.01.22; peer-reviewed by R Kammler, S McLennan, T Lefèvre; comments to author 23.03.22; revised version received 02.05.22; accepted 17.08.22; published 29.09.22.*

*Please cite as:*

Franchini F, Kusejko K, Marzolini C, Tellenbach C, Rossi S, Stampf S, Koller M, Stoyanov J, Möller B, Leichtle AB

*Collaborative Challenges of Multi-Cohort Projects in Pharmacogenetics—Why Time Is Essential for Meaningful Collaborations*

*JMIR Form Res* 2022;6(9):e36759

URL: <https://formative.jmir.org/2022/9/e36759>

doi: [10.2196/36759](https://doi.org/10.2196/36759)

PMID: [35976179](https://pubmed.ncbi.nlm.nih.gov/35976179/)

©Filippo Franchini, Katharina Kusejko, Catia Marzolini, Christoph Tellenbach, Simona Rossi, Susanne Stampf, Michael Koller, Jivko Stoyanov, Burkhard Möller, Alexander Benedikt Leichtle. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 29.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Exploring Motivations for COVID-19 Vaccination Among Black Young Adults in 3 Southern US States: Cross-sectional Study

Marie CD Stoner<sup>1</sup>, MPH, PhD; Erica N Browne<sup>1</sup>, BSc, MSc; David Tweedy<sup>1</sup>, BSc; Audrey E Pettifor<sup>2</sup>, MPH, PhD; Allysha C Maragh-Bass<sup>3</sup>, MPH, PhD; Christina Toval<sup>2</sup>, MPH; Elizabeth E Tolley<sup>3</sup>, MA, PhD; Maria Leonora G Comello<sup>2</sup>, MSc, PhD; Kathryn E Muessig<sup>2</sup>, PhD; Henna Budhwani<sup>4</sup>, MPH, PhD; Lisa B Hightow-Weidman<sup>2</sup>, MD, MPH

<sup>1</sup>RTI International, Berkeley, CA, United States

<sup>2</sup>University of North Carolina, Chapel Hill, NC, United States

<sup>3</sup>FHI 360, Durham, NC, United States

<sup>4</sup>Florida State University, Tallahassee, FL, United States

**Corresponding Author:**

Marie CD Stoner, MPH, PhD

RTI International

2150 Shattuck Avenue

Berkeley, CA, 94704

United States

Phone: 1 510 665 8285 ext 34285

Email: [mcstoner@rti.org](mailto:mcstoner@rti.org)

## Abstract

**Background:** Few studies have focused on attitudes toward COVID-19 vaccination among Black or African American young adults (BYA) in the Southern United States, despite high levels of infection in this population.

**Objective:** To understand this gap, we conducted an online survey to explore beliefs and experiences related to COVID-19 vaccination among BYA (aged 18-29 years) in 3 southern states.

**Methods:** We recruited 150 BYA to participate in an online survey as formative research for an intervention to address vaccine hesitancy in Alabama, Georgia, and North Carolina from September 22, 2021, to November 18, 2021. Participants were recruited through social media ads on Facebook, Twitter, Instagram, and YouTube. Additionally, we distributed information about the survey through organizations working with BYA in Alabama, Georgia, and North Carolina; our community partners; and network collaborations. We used measures that had been used and were previously validated in prior surveys, adapting them to the context of this study.

**Results:** Roughly 28 (19%) of the participants had not received any doses of the COVID-19 vaccine. Half of the unvaccinated respondents (n=14, 50%) reported they wanted to wait longer before getting vaccinated. Motivators to get vaccinated were similar between unvaccinated and vaccinated respondents (eg, if required, to protect the health of others), but the main motivator for those vaccinated was to protect one's own health. Among unvaccinated individuals, reasons for not receiving the COVID-19 vaccine included concern about vaccine side effects (n=15, 54%) and mistrust of vaccine safety (n=13, 46%), of effectiveness (n=12, 43%), and of the government's involvement with vaccines (n=12, 43%). Experiences of discrimination (n=60, 40%) and mistrust of vaccines (n=54, 36%) were common overall. Among all respondents, those who said they would be motivated to get vaccinated if it was required for school, work, or travel were more likely to endorse negative beliefs about vaccines compared to those motivated for other reasons.

**Conclusions:** Mistrust in COVID-19 vaccine safety and efficacy is common among BYA in the Southern United States, irrespective of vaccination status. Other motivators, such as safety of family and community and vaccination requirements, may be able to tip the scales toward a decision to be vaccinated among those who are initially hesitant. However, it is unclear how vaccine requirements among BYA in the South affect trust in the government or health care in the long term. Interventions that include BYA in vaccination messaging and programs may more proactively build feelings of trust and combat misinformation.

(*JMIR Form Res* 2022;6(9):e39144) doi:[10.2196/39144](https://doi.org/10.2196/39144)

## KEYWORDS

COVID-19; COVID-19 vaccination; young people; vaccination motivations; vaccination beliefs; online survey; health disparity; minority population; vaccine hesitancy; misinformation; vaccine safety

## Introduction

Young adults (18-29 years) are a priority population among which to increase COVID-19 vaccine uptake because they have high rates of asymptomatic infection and transmission [1-3]. For example, a study of 767 hotspot counties early in the pandemic in 2020 reported that percentage positivity increased earliest in younger persons (aged 18-24 years), followed by several weeks of increasing percentage positivity among older age groups, particularly in hot spots within the Southern and Western United States [1]. These findings corroborate patterns seen in another study from the Southern United States, where increased percentage positivity among young adults preceded increases among those aged  $\geq 60$  years by 4-15 days [4]. Given that many young adults live in multigenerational households with older loved ones, they are also potentially a source of COVID-19 transmission to older adults, who are more likely to be vulnerable to COVID-19 irrespective of race [5]. Roughly, a third of young adults lived in multigenerational households in 2021, and Black or African American young adults (BYA) were more likely than White young adults to live in these multigenerational households [6]. Although young people may be more likely to transmit COVID-19 than other demographics, they are also potential change agents to promote safer behaviors within their families and communities. Young adults who identify as Black or African American (referred to as Black throughout this paper) have critical social capital to influence health behaviors in others, including in adult family and community members who are at high risk for COVID-19 hospitalization and mortality [7]. Therefore, BYA should be prioritized in efforts to increase COVID-19 vaccination because they make an important contribution to sustaining community transmission and can serve as change agents in their communities.

Multiple vaccines have been proven to be efficacious in preventing hospitalization and severe COVID-19 infection and are now widely available [8,9]. Despite the high efficacy, the rapid development and approval of these vaccines have resulted in public concern around vaccine safety, and many remain hesitant toward vaccination [10]. Concern around vaccines is particularly salient in Black communities with deeply rooted medical mistrust due, in part, to well-known ethical violations, such as Henrietta Lacks and the Tuskegee study and structures, policies, and practices rooted in structural racism that create health inequities in many Black communities [10-14]. Several studies have identified that vaccine hesitancy varies by age and race, with Black communities having higher rates of vaccine hesitancy and younger age being associated with greater hesitancy across race/ethnicity [10,15-17]. However, few studies focus specifically on BYA in the Southern United States who are an important population among which to increase vaccination, including both completion of the primary vaccine series and receipt of booster shots. As of March 28, 2022, the percentage with at least 1 dose of the vaccine was 65% in

Georgia, 62% in Alabama, and 83% in North Carolina, lower in 2 of the 3 states than the US national average of 77% [18]. We conducted an online survey with BYA (aged 18-29 years) in Georgia, Alabama, and North Carolina to describe rates of COVID-19 vaccination acceptance and known correlates (their experiences with health care, beliefs around vaccines in general and with the COVID-19 vaccine, and demographic factors). This survey was designed as a formative assessment to inform content for a digital health intervention to address COVID-19 vaccination hesitancy among BYA in these states.

## Methods

### Study Population and Recruitment

We recruited 150 BYA to participate in an online survey related to COVID-19 vaccination in Alabama, Georgia, and North Carolina in the Southern United States (Budhwani et al, unpublished data, 2022). Eligibility criteria for the online survey included (1) 18-29 years of age, (2) identified as African American or Black, (3) English proficient, (4) access to a personal smartphone, and (5) resident of Georgia, Alabama, or North Carolina. We aimed to recruit an approximately equal number of respondents from each state: 44 (29%) respondents from Alabama, 50 (33%) from Georgia, and 56 (38%) from North Carolina. Participants completed a 1-time online survey through the Qualtrics survey platform.

Participants were recruited through social media ads on Facebook, Twitter, Instagram, and YouTube. Additionally, we distributed information about the survey through organizations in the study areas working with BYA, our community partners, and other network collaborations. BYA interested in participating in the study clicked a link within an email notification or via social media and were redirected to a study website, where they completed eligibility screening. All potential participants completed an online screening survey via Qualtrics to obtain consent to be screened and to verify inclusion criteria. The online screening survey was about 5 minutes in length and included a script that was read by the participants that explained the purpose of screening and the general purpose of the study and clarified that if they were eligible, they would be invited to participate in the study. We used established stringent verification procedures utilized in multiple prior studies to address fraud (eg, assessment for duplicate entries and suspicious response patterns, confirmation of the internet protocol [IP] address, zip code, and latitude and longitude) [19-21]. Eligible BYA were then directed to an informed consent page. Participants agreed to participate by clicking forward and providing online electronic consent (e-consent) or declined by exiting the website. The survey took approximately 30-45 minutes to complete. All respondents received a US \$20 incentive for participation and the ability to participate in a lottery, where they were eligible to receive a larger prize.

The survey was conducted from September to November 2021. At this time during the COVID-19 pandemic, vaccines were

available for all respondents who were aged 18-29 years. The Food and Drug Administration (FDA) issued an Emergency Use Authorization (EUA) for the Pfizer-BioNTech COVID-19 vaccine on December 11, 2020; for the Moderna vaccine on December 28, 2020; and for the Janssen (Johnson & Johnson) COVID-19 vaccine on February 27, 2021 [22]. Before the study period and on September 22, 2021, the FDA authorized a booster dose of the Pfizer-BioNTech COVID-19 vaccine for individuals aged 65 years and older, aged 18-64 years at high risk of severe COVID-19, and aged 18-64 years who had institutional or occupational exposure to SARS-CoV-2 [22]. On November 29, 2021, just after the survey period, the Centers for Disease Control and Prevention (CDC) expanded its recommendation that all individuals aged 18 years and older get a COVID-19 booster [22].

### Ethical Considerations

The study was overseen by the University of North Carolina (UNC), Chapel Hill (Approval #21-1746), with a reliance agreement from the Institutional Review Boards at the University of Alabama, RTI International, and FHI 360.

### Measures

Survey measures were selected according to constructs of social cognitive theory (SCT) to understand behavioral, cognitive (eg, knowledge), and environmental (eg, social influences by parents, peers, and community) influences motivating COVID-19 vaccination. We used measures that had been used and were previously validated in prior surveys, adapting them to the context of this study. To make these domains and constructs more specific to COVID-19 vaccination, we referred to the National Institute on Minority Health and Health Disparities (NIMHD) research framework, which is informed by the World Health Organization (WHO) Strategic Advisory Group of Experts (SAGE) [23]. We adapted the NIMHD framework to reflect determinants of vaccine hesitancy for BYA within different levels of influence (eg, individual, interpersonal, community) [24]. We examined individual, interpersonal, and community influences on vaccination as these different sources of social influence have all been associated with health behaviors in young adults. Vaccine beliefs were measured using 10 items on a 5-point Likert scale adapted from a validated scale developed by Shapiro et al [25]. Measures of vaccine confidence were based on the CDC COVID-19 Vaccine Confidence Rapid Community Assessment Tool, which is validated and has been previously used [26]. Questions related to reasons and motivations for vaccination were adapted from the RADX-up Common Data Elements (CDEs) set of validated measures [27]. When asked about their motivations, respondents were asked to select their main motivation for being or not being vaccinated from a preset list and to choose all that apply [27].

### Statistical Analysis

We descriptively examined characteristics, including participant demographics and knowledge, attitudes, and beliefs regarding vaccination, in general and about the COVID-19 vaccine, as well as prevention practices in the overall sample. We used frequencies and percentages for categorical variables and medians with the IQR for continuous variables. We tested for difference in vaccination by state using a chi-square test. We also examined motivations and barriers among those fully vaccinated (2 doses of a 2-dose vaccine, or 1 dose of a 1-dose vaccination) versus those not vaccinated and grouped vaccine hesitancy characteristics into their corresponding SCT domains/constructs to assess how these components influence vaccination behaviors. Although some measures were items from scales, all items were examined individually rather than as a score in order to understand answers to specific responses. The full survey is available upon request. In the results, we focus on environmental determinants of vaccination, which were a large contributor to overall attitudes about vaccination.

## Results

### Demographic Characteristics

Among 150 BYA who completed the survey between from September to November 2021, the median age was 23 (IQR 20-26) years, 88 (59%) identified as cisgender women, 41 (27%) as cisgender men, 4 (3%) as transgender, and 18 (12%) as nonbinary, genderqueer, gender nonconforming, or gender fluid (Table 1). One-third were employed as essential workers (n=49, 33%), 44 (29%) lived in a household with any children under 18 years old, and 61 (61%) either had a chronic condition themselves or currently lived with someone with a chronic health condition. Most respondents (n=123, 82%) reported ever having at least 1 negative experience with a health care provider, where a provider did not believe they were telling the truth, assumed something without asking, did not listen to them, or suggested they were personally to blame for a health problem. Half (n=69, 49%) thought they were treated this way specifically because of their race.

Approximately 112 (75%) of the respondents had received 2 doses of a messenger RNA (mRNA) vaccine or 1 dose of the Johnson & Johnson COVID-19 vaccine, 10 (7%) received 1 dose of a 2-dose vaccine, and 28 (19%) had not received a vaccine. Vaccination was similar across states (n=49, 87% in North Carolina; n=38, 76% in Georgia; n=35, 80% in Alabama;  $P=.30$ ). Among those who were fully vaccinated, less than half (n=71, 47%) planned to get a booster shot; however, this was before the CDC strengthened its recommendation on November 29, 2021, for all individuals over 18 years old to receive the booster [22]. Less than 15% had ever been diagnosed with COVID-19 (n=18, 12%), but 79 (53%) knew someone who had been hospitalized or died as a result of having COVID-19.



**Table 1.** Demographics of 150 BYA<sup>a</sup> who participated in the survey (September-November 2021).

Characteristics	Participants
Age (years), median (IQR)	23 (20-26)
Female sex, n (%)	103 (69)
<b>Gender identity, n (%)</b>	
Cisgender woman	88 (59)
Cisgender man	41 (27)
Transgender	4 (3)
Nonbinary, genderqueer, gender nonconforming, or gender fluid	18 (12)
<b>Sexual orientation, n (%)</b>	
Gay or lesbian	10 (7)
Straight	87 (58)
Bisexual	35 (23)
Queer, pansexual, asexual, or other	28 (19)
<b>Highest level of education completed, n (%)</b>	
Less than high school	2 (1)
High school	103 (69)
Bachelor's degree or higher	45 (30)
Employed as an essential worker, n (%)	49 (33)
Have health insurance, n (%)	106 (71)
Household size, median (IQR)	2 (1-4)
Children <18 years old in the household, n (%)	44 (29)
Self or household member with a chronic health condition, n (%)	91 (61)
Smoke tobacco, n (%)	22 (15)
<b>State of residence, n (%)</b>	
Alabama	44 (29)
Georgia	50 (33)
North Carolina	56 (37)
<b>COVID-19 vaccination and diagnosis status, n (%)</b>	
Received both doses or one dose of the Johnson & Johnson	112 (75)
Received one dose of a two-dose vaccine, intend to get second dose	10 (7)
Not received the vaccine	28 (19)
Planning to get a booster	71 (47)
Ever diagnosed with COVID-19	18 (12)

<sup>a</sup>BYA: Black or African American young adults.

## Motivations for Vaccination

Of those BYA who were fully vaccinated (n=122, 81%), the main reason for accepting the vaccine was to protect their own health (n=56, 46%); to protect the health of family/friends, coworkers, or community members (n=33, 27%); and to resume work, school, social activities, or travel (n=26, 21%); see [Table 2](#). When able to choose more than 1 motivation, the mean number of reasons selected for receiving the COVID-19 vaccine was 4.1 (range 1-8). About half (n=56, 46%) reported that most or all family members had been vaccinated, 69 (57%) reported

that most or all close friends had been vaccinated, and 87 (71%) reported that community leaders supported vaccination. Most were confident that the COVID-19 vaccines are safe (n=109, 85%) and effective at preventing severe illness (n=109, 89%), but 37 (30%) had heard information about the COVID-19 vaccine that they could not determine was true. In addition, 34 fully vaccinated respondents (28%) had little to no trust in the government and public health agencies' vaccine recommendations.

Among those who were unvaccinated (n=28, 19%), the main motivators to get vaccinated were if it was required for school,

work, or travel (n=10, 36%), to protect the health of family/friends (n=7, 25%), and to protect one's own health (n=3, 11%); 4 (14%) participants did not know what would motivate them (Table 3). Less than a third were confident that the COVID-19 vaccines are safe (n=8, 29%) and effective (n=6, 21%). The main reasons unvaccinated BYA noted not receiving the COVID-19 vaccine included concern about vaccine side effects (n=6, 21%), not trusting that the vaccine is safe (n=7, 25%), not knowing enough about how well a COVID-19 vaccine

works (n=4, 14%), waiting until more people have gotten it (n=4, 14%), and not trusting the government (n=3, 11%). When able to choose more than 1 main motivator, the mean number of reasons selected for not receiving the COVID-19 vaccine was 2.3 (range 0-6). Half of the unvaccinated sample (n=14, 50%) reported wanting to wait to get vaccinated until it had been available for a while to see how it is working for other people; 4 (14%) participants said they would definitely not get the vaccine.

**Table 2.** Motivations for receiving the COVID-19 vaccine.

Motivations	Participants (N=122)
<b>Main reason chose to get the COVID-19 vaccine, n (%)</b>	
To protect my health	56 (46)
To protect the health of family/friends, coworkers or my community	33 (27)
To resume work, school, social activities, or travel	26 (21)
Because others I trust encouraged me to get it	5 (4)
Other	2 (3)
Mean number of reasons selected for COVID-19 vaccination when able to choose more than 1, mean (range)	4.1 (1-8)

**Table 3.** Reasons for not getting the COVID-19 vaccine.

Reasons	Participants (N=28)
<b>Main reason for not getting the COVID-19 vaccine, n (%)</b>	
Do not trust that the vaccine is safe	7 (25)
Concerned about side effects from the vaccine	6 (21)
Waiting until more people have gotten it	4 (14)
Do not know enough about how well a COVID-19 vaccine works	4 (14)
Do not trust the government	3 (11)
Just have not gotten around to getting it	2 (7)
Other	2 (7)
Mean number of reasons selected for not getting COVID-19 vaccination when able to choose more than 1, mean (range)	2.3 (0-6)
<b>Main reason that would motivate to get the COVID-19 vaccine, n (%)</b>	
To protect my health	3 (11)
To protect the health of family/friends, coworkers, or community	7 (25)
To resume work, school, social activities, or travel	10 (36)
Because others I trust encouraged me to get it	1 (4)
Other	3 (11)
Do not know	4 (14)

### Attitudes Toward Vaccines

Overall, nearly all respondents (n=138, 92%) believed that, in general, vaccines are useful and effective, and many (n=90, 60%) reported prior lifetime receipt of the flu vaccine (Table 4). However, there was a high level of overall mistrust in vaccines; the participants believed that people are deceived about how well vaccines work to prevent illness (n=54, 36%) and about vaccine safety (n=55, 37%). Approximately one-third thought pharmaceutical companies cover up the dangers of vaccines (n=47, 31%) and that the government is experimenting

with vaccines on the Black community (n=43, 29%); see Table 4.

Vaccine beliefs differed by the underlying motivation to receive the COVID-19 vaccine. Those who said they were or would be motivated to get vaccinated if it was required for school, work, or travel were more likely to endorse all negative beliefs about vaccination compared to those motivated for other reasons (Table 5). For example, 21 (58%) of those motivated by vaccine requirements agreed that the government is experimenting with vaccines on the Black community compared to 10 (25%) of

those motivated to protect health of others and 11 (19%) of those motivated to protect their own health. Most (n=21, 58%) of those motivated by vaccine requirements agreed that people are deceived about vaccine effectiveness compared to 9 (23%) of those motivated to protect health of others and 15 (25%) of those motivated to protect their own health.

**Table 4.** Beliefs about vaccination and experiences with health care.

Beliefs and experiences	Participants (N=150), n (%)
<b>Social network and COVID-19</b>	
Know anyone who has been vaccinated	133 (89)
Most or all family members vaccinated	61 (41)
Most or all close friends vaccinated	76 (51)
Leaders (religious, political, teachers, health care workers) in your community support getting vaccinated	103 (69)
Know someone personally who has been hospitalized or died because of COVID-19	79 (53)
<b>COVID-19 vaccine confidence</b>	
Confident currently available COVID-19 vaccines in the United States are effective at preventing severe illness sick or hospitalization	115 (77)
Confident COVID-19 vaccines currently available in the United States are safe	112 (75)
Know where to get accurate, timely information about COVID-19 vaccines	110 (73)
I have heard information about COVID-19 vaccines that I cannot determine is true or false	50 (33)
Trust the government and public health agencies that recommend getting a COVID-19 vaccine	94 (62)
<b>Negative experiences with health care: Have you ever felt that a doctor or health care provider:</b>	
Did not believe you were telling the truth?	81 (54)
Refused to order a test or treatment you thought you needed?	69 (46)
Suggested you were personally to blame for a health problem you were experiencing?	67 (45)
Assumed something about you without asking?	83 (55)
Talked down to you or did not treat you with respect?	73 (49)
Did not listen to what you had to say?	81(54)
Ever experience racial discrimination from health care provider?	60 (40)
<b>Vaccine beliefs (agree or strongly agree)</b>	
Vaccine safety data are often fabricated.	33 (22)
Pharmaceutical (drug) companies cover up the dangers of vaccines.	47 (31)
People are deceived about how well vaccines work to prevent illness and death.	54 (36)
People are deceived about vaccine safety.	55 (37)
The government covers up the link between vaccines and autism.	22 (15)
The government is experimenting with vaccines on the Black community.	43 (29)
Vaccines have many known harmful side effects.	50 (33)
Vaccines may lead to illness and death.	43 (29)
Vaccines provide important benefits to society.	133 (89)
Vaccines are useful and effective.	138 (92)
<b>Vaccine experience</b>	
Ever received the flu vaccine	90 (60)
Received the flu vaccine in the past 12 months	44 (29)

**Table 5.** Vaccine beliefs by motivations for receiving the COVID-19 vaccine.

Vaccine beliefs	COVID-19 vaccination motivation, n (%)				Total
	To protect my health	To protect the health of others	Required	Other or do not know <sup>a</sup>	
Total	59 (100)	40 (100)	36 (100)	15 (100)	150 (100)
<b>Vaccine safety data are often fabricated.</b>					
Agree	8 (14)	7 (18)	12 (33)	6 (40)	33 (22)
Disagree or do not know	51 (86)	33 (83)	24 (67)	9 (60)	117 (78)
<b>Pharmaceutical (drug) companies cover up the dangers of vaccines.</b>					
Agree	14 (24)	6 (15)	17 (47)	10 (67)	47 (31)
Disagree or do not know	45 (76)	34 (85)	19 (53)	5 (33)	103 (69)
<b>People are deceived about how well vaccines work to prevent illness and death.</b>					
Agree	15 (25)	9 (23)	21 (58)	9 (60)	54 (36)
Disagree or do not know	44 (75)	31 (78)	15 (42)	6 (40)	96 (64)
<b>People are deceived about vaccine safety.</b>					
Agree	18 (31)	8 (20)	20 (56)	9 (60)	55 (37)
Disagree or do not know	41 (70)	32 (80)	16 (44)	6 (40)	95 (63)
<b>The government covers up the link between vaccines and autism.</b>					
Agree	5 (9)	5 (13)	6 (17)	6 (40)	22 (15)
Disagree or do not know	54 (92)	35 (88)	30 (83)	9 (60)	128 (85)
<b>The government is experimenting with vaccines on the Black community.</b>					
Agree	11 (19)	10 (25)	17 (47)	5 (33)	43 (29)
Disagree or do not know	48 (81)	30 (75)	19 (53)	10 (67)	107 (71)

<sup>a</sup>Because others I trust encouraged me to get it" (n=6, 40%), "other" (n=5, 33%), and "do not know" (n=4, 27%).

## Discussion

### Principal Findings

We conducted an online survey to describe COVID-19 vaccine acceptance and correlates among young Black adults in 3 southern states experiencing COVID-19 inequities. We found that approximately 14% of respondents had not received any doses of the COVID-19 vaccine as of November 2021. Most unvaccinated respondents reported that they were waiting to get vaccinated until it had been available for a while. Both vaccinated and unvaccinated respondents described similar motivations for vaccination, including to protect their own health, to protect the health of their family and communities, and requirements to resume work, school, travel, or social activities. Mistrust of vaccines was common even among the vaccinated population and was a barrier among the unvaccinated. Among all respondents, those who said they were or would be motivated to get vaccinated if it was required for school, work, or travel were more likely to endorse negative beliefs about vaccination compared to those motivated for health reasons. Overall, our findings show that strategies to increase vaccination among BYA should emphasize the health of the family and community, while also continuing to build trust in vaccination overall and specifically around the safety and efficacy of the COVID-19 vaccine. Strategies that use vaccination mandates

should consider and address how these requirements may also affect trust in vaccination. These findings are being used to inform the content for a smartphone application to address vaccination BYA in Georgia, Alabama, and North Carolina [24] (Budhwani et al, unpublished data, 2022).

### Comparison With Prior Work

Despite the relatively young age of our sample, experiences of discrimination by a health care provider and mistrust of vaccines and the government's involvement with vaccines were common. Among those unvaccinated, the main concerns about receiving the COVID-19 vaccine were side effects, trust of the government, trust in vaccines in general and specifically of the safety and efficacy of the COVID-19 vaccine, and not having enough information about how well the COVID-19 vaccine works. Our findings are similar to a systematic review of 13 studies that found that major predictors of vaccine hesitancy in Black and Hispanic people were medical mistrust and a history of racial discrimination; exposure to myths and misinformation; beliefs about vaccines and past vaccine compliance; and concerns about the safety, efficacy, and side effects of the COVID-19 vaccines [10]. Concerns about safety and efficacy have been shown to be a major barrier to vaccination across populations and countries [28]. Additional studies among college populations have also shown that a common reason for unwillingness to receive the vaccine are the belief that the

vaccine approval process was rushed [15], as well as mistrust of the health care system or government held by Black individuals, as described earlier [29]. Our findings indicate that mistrust was also common among vaccinated BYA, suggesting the potential for other motivations to build trust and facilitate vaccine uptake, such as trusted sources to combat misinformation and inclusion of BYA in vaccination messaging and programs, given their potential as change agents.

A review of interventions to increase vaccination for any indication found that aside from provider recommendations, the most effective interventions to increase vaccination have been those that target direct behavior change through the use of policies and practices to increase vaccination (eg, mandates, incentives, transportation) without changing what people think or feel (eg, vaccine confidence or perceived risk) [30-32]. Evidence from randomized trials has also shown that interventions to build on intentions (eg, reminders, prompts, and reducing logistical barriers) and shape behavior (eg, incentives, sanctions, or requirements, such as work and school vaccination mandates) have been effective at increasing vaccination [30-32]. In our survey, we found that many respondents were waiting to eventually receive the vaccine and might benefit from strategies to facilitate vaccination, although vaccines had already been approved for almost a year at the time of the survey [22]. In addition, we found that almost a third of those who were vaccinated and unvaccinated reported that their main motivation for vaccination was to resume work, school, travel, or other social activities. Yet, those who said they were or would be motivated because of vaccine requirements were more likely to endorse negative beliefs about vaccination compared to those motivated by other reasons. Although requirements may be an important strategy and can result in greater vaccine uptake, it still leaves a gap in building trust in vaccines and does nothing to address or may reinforce justified medical skepticism within the Black community. Effective intervention strategies to address mistrust have included communicating the importance of vaccination through trusted channels (eg, Black physicians and clinicians), validating the real history- and experience-based reasons people may be hesitant, and addressing racism embedded within the health care system [33]. Interventions are critically needed to build trust and address medical mistrust, including through meaningful involvement of BYA in policies and programming for this age group.

Medical mistrust and a history of abuse and racism from the medical and research establishment mean that increasing vaccine uptake and trust in the safety and efficacy of COVID-19 vaccines requires messaging with involvement from trusted sources in the Black community and that the role of social support in encouraging vaccination is key. Our findings largely resonate with those of a prior study by Balasuriya et al [34] that used qualitative interviews with 72 respondents to explore COVID-19 vaccine access and acceptance among Black and Latinx communities in March 2021. The study identified 3 major themes that represent facilitators of and barriers to COVID-19 vaccination: (1) pervasive mistreatment of Black and Latinx communities and associated distrust; (2) informing trust via trusted messengers and messages, choice, social support, and

diversity; and (3) addressing structural barriers to vaccination access. The first theme is related to pervasive mistreatment of Black and Latinx communities both historically and throughout the pandemic and aligns with the high levels of mistrust that were observed in our sample and high reported experiences of discrimination in health care. Within the second theme, respondents described feelings of trust in the COVID-19 vaccine that were built through receiving information from trusted messengers, consistent and transparent messaging, social support, and seeing diversity at the vaccination site. Accordingly, our results also show that despite mistrust even among vaccinated individuals, family and community concern was a motivator to receive the vaccine. Additionally, questions and skepticism about vaccination can be protective and efforts are needed to ensure that institutions are trustworthy, transparent, and engaged with communities throughout the vaccine rollout and other public health efforts [35].

### Limitations

Our study used a convenience sample of young adults from North Carolina, Georgia, and Alabama and may not be representative of the larger population of BYA in these states. We observed that 86% of our sample was vaccinated, a rate much higher than national- and state-level data [36,37]. As of November 1, 2021, 59.3% of adults over the age of 18 years were vaccinated in Georgia, 54.8% in Alabama, and 63.4% in North Carolina [37]. The higher percentage vaccinated in our study may be because unvaccinated BYA are less likely to want to participate in research or because the survey was not distributed as widely through the social networks of unvaccinated individuals. There was also an overrepresentation of those assigned the female sex at birth (69%) in the sample. In addition, our sample included only 150 BYA, with 28 unvaccinated individuals. Given the small sample size and biased sample, we had limited power and ability to make statements about this population or specific comparisons between those vaccinated or unvaccinated. Lastly, the science and evidence have been evolving quickly on COVID-19, and this survey was conducted in September-November 2021, before the CDC expanded its recommendation on November 29, 2021, that all individuals ages 18 years and older get a COVID-19 booster [22]. Therefore, we are unable to make conclusions about boosters in this population, although more recent evidence suggests the percentage of those who have received a booster shot is lower in those aged 18-29 years compared to other age groups [36]. Future studies are needed to determine strategies to increase booster shots, particularly as concerns about vaccination were common among both vaccinated and unvaccinated groups in our survey and may lead to a lower likelihood of boosters. We also did not ask about COVID-19 testing since 2020. Therefore, the percentage who were ever diagnosed with COVID-19 (12%) is likely an underestimate of the percentage of the population who had truly been infected.

### Conclusion

In a convenience sample of 150 BYA (aged 18-29 years) in Georgia, Alabama, and North Carolina, most respondents had received the COVID-19 vaccine or reported wanting to wait to get vaccinated until it had been available for a while to see how

it is working for other people. Reporting of mistrust of vaccines, particularly of vaccine safety and efficacy, was common and was noted as a barrier to receiving a vaccine among those who were unvaccinated and was also pervasive among those who were vaccinated. Protecting the health of the family and community, protecting individual health, and vaccine requirements were motivators for vaccination among both vaccinated and unvaccinated individuals. Yet, those who said they were or would be motivated because of vaccine requirements were more likely to endorse negative beliefs about vaccination compared to those motivated by other reasons. Strategies such as vaccine requirements may be able to tip the scales toward a decision to get vaccinated among those who are

initially hesitant—at least in the short term. However, it is unclear how vaccine requirements among BYA in the South will affect trust in the government or health care in the long term. Our study suggests that whether vaccine accepting or hesitant, interventions are critically needed to build trust in vaccines and address justified medical skepticism within the Black community. Interventions to address COVID-19 vaccination in BYA in the South should focus on building feelings of trust in the COVID-19 vaccine through receiving information from trusted messengers, consistent and transparent messaging, and social support, with meaningful involvement from BYA in COVID-19 policies and programming for this age group.

---

## Acknowledgments

We would like to thank the respondents and the study team. This study was funded by a grant from the National Institute on Minority Health and Health Disparities (NIMHD; #R01MD016834; #NCT05490329).

---

## Conflicts of Interest

None declared.

---

## References

1. Oster AM, Caruso E, DeVies J, Hartnett KP, Boehmer TK. Transmission dynamics by age group in COVID-19 hotspot counties - United States, April-September 2020. *MMWR Morb Mortal Wkly Rep* 2020 Oct 16;69(41):1494-1496 [FREE Full text] [doi: [10.15585/mmwr.mm6941e1](https://doi.org/10.15585/mmwr.mm6941e1)] [Medline: [33056949](https://pubmed.ncbi.nlm.nih.gov/33056949/)]
2. Salvatore PP, Sula E, Coyle JP, Caruso E, Smith AR, Levine RS, et al. Recent increase in COVID-19 cases reported among adults aged 18-22 years - United States, May 31-September 5, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Oct 02;69(39):1419-1424 [FREE Full text] [doi: [10.15585/mmwr.mm6939e4](https://doi.org/10.15585/mmwr.mm6939e4)] [Medline: [33006586](https://pubmed.ncbi.nlm.nih.gov/33006586/)]
3. Huang L, Zhang X, Zhang X, Wei Z, Zhang L, Xu J, et al. Rapid asymptomatic transmission of COVID-19 during the incubation period demonstrating strong infectivity in a cluster of youngsters aged 16-23 years outside Wuhan and characteristics of young patients with COVID-19: a prospective contact-tracing study. In: *J Infect. Rapid asymptomatic transmission of COVID-19 during the incubation period demonstrating strong infectivity in a cluster of youngsters aged 16-23 years outside Wuhan and characteristics of young patients with COVID-19: A prospective contact-tracing study. J Infect Internet J Infect*; Jun 2020:e1-e13.
4. Boehmer TK, DeVies J, Caruso E, van Santen KL, Tang S, Black CL, et al. Changing age distribution of the COVID-19 pandemic - United States, May-August 2020. *MMWR Morb Mortal Wkly Rep* 2020 Oct 02;69(39):1404-1409 [FREE Full text] [doi: [10.15585/mmwr.mm6939e1](https://doi.org/10.15585/mmwr.mm6939e1)] [Medline: [33001872](https://pubmed.ncbi.nlm.nih.gov/33001872/)]
5. Aparicio Fenoll A, Grossbard S. Intergenerational residence patterns and Covid-19 fatalities in the EU and the US. *Econ Hum Biol* 2020 Dec;39:100934 [FREE Full text] [doi: [10.1016/j.ehb.2020.100934](https://doi.org/10.1016/j.ehb.2020.100934)] [Medline: [33160264](https://pubmed.ncbi.nlm.nih.gov/33160264/)]
6. PEW Research Center. Demographics of Multigenerational Households. 2022. URL: <https://www.pewresearch.org/social-trends/2022/03/24/the-demographics-of-multigenerational-households/> [accessed 2022-08-18]
7. Ginwright S. Black youth activism and the role of critical social capital in black community organizations. *Am Behav Sci* 2016 Jul 27;51(3):403-418. [doi: [10.1177/0002764207306068](https://doi.org/10.1177/0002764207306068)]
8. Tartof SY, Slezak JM, Fischer H, Hong V, Ackerson BK, Ranasinghe ON, et al. Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. *Lancet* 2021 Oct;398(10309):1407-1416. [doi: [10.1016/s0140-6736\(21\)02183-8](https://doi.org/10.1016/s0140-6736(21)02183-8)]
9. Pilishvili T, Fleming-Dutra KE, Farrar JL, Gierke R, Mohr NM, Talan DA, Vaccine Effectiveness Among Healthcare Personnel Study Team. Interim estimates of vaccine effectiveness of Pfizer-BioNTech and Moderna COVID-19 vaccines among health care personnel — 33 U.S. Sites, January–March 2021. *MMWR Morb Mortal Wkly Rep* 2021 May 21;70(20):753-758 [FREE Full text] [doi: [10.15585/mmwr.mm7020e2](https://doi.org/10.15585/mmwr.mm7020e2)] [Medline: [34014909](https://pubmed.ncbi.nlm.nih.gov/34014909/)]
10. Khubchandani J, Macias Y. COVID-19 vaccination hesitancy in Hispanics and African-Americans: a review and recommendations for practice. *Brain Behav Immun Health* 2021 Aug;15:100277 [FREE Full text] [doi: [10.1016/j.bbih.2021.100277](https://doi.org/10.1016/j.bbih.2021.100277)] [Medline: [34036287](https://pubmed.ncbi.nlm.nih.gov/34036287/)]
11. Maragh-Bass A, Shackelford L. The Full Picture: Accurately Framing COVID-19 Vaccine “Hesitancy” among Black Americans. 2021. URL: <https://degrees.fhi360.org/2021/04/the-full-picture-accurately-framing-covid-19-vaccine-hesitancy-among-black-americans/> [accessed 2022-08-18]

12. Bogart L, Ojikutu B, Tyagi K, Klein D, Mutchler M, Dong L, et al. COVID-19 related medical mistrust, health impacts, potential vaccine hesitancy among black Americans living with HIV. *J Acquir Immune Defic Syndr* 2021 Feb;86(2):200-207. [doi: [10.1097/qai.0000000000002570](https://doi.org/10.1097/qai.0000000000002570)]
13. Manning KD. More than medical mistrust. *Lancet* 2020 Nov;396(10261):1481-1482. [doi: [10.1016/s0140-6736\(20\)32286-8](https://doi.org/10.1016/s0140-6736(20)32286-8)]
14. Corbie-Smith G. Vaccine hesitancy is a scapegoat for structural racism. *JAMA Health Forum* 2021 Mar 25;2(3):e210434. [doi: [10.1001/jamahealthforum.2021.0434](https://doi.org/10.1001/jamahealthforum.2021.0434)]
15. Gurley S, Bennett B, Sullivan P, Kiley M, Linde J, Szczerbacki D, et al. COVID-19 vaccine perceptions, intentions, and uptake among young adults in the United States: prospective college-based cohort study. *JMIR Public Health Surveill* 2021 Dec 15;7(12):e33739 [FREE Full text] [doi: [10.2196/33739](https://doi.org/10.2196/33739)] [Medline: [34847054](https://pubmed.ncbi.nlm.nih.gov/34847054/)]
16. Moore JX, Gilbert KL, Lively KL, Laurent C, Chawla R, Li C, et al. Correlates of COVID-19 vaccine hesitancy among a community sample of African Americans living in the Southern United States. *Vaccines (Basel)* 2021 Aug 08;9(8):879 [FREE Full text] [doi: [10.3390/vaccines9080879](https://doi.org/10.3390/vaccines9080879)] [Medline: [34452004](https://pubmed.ncbi.nlm.nih.gov/34452004/)]
17. Kricorian K, Turner K. COVID-19 vaccine acceptance and beliefs among Black and Hispanic Americans. *PLoS One* 2021 Aug 24;16(8):e0256122 [FREE Full text] [doi: [10.1371/journal.pone.0256122](https://doi.org/10.1371/journal.pone.0256122)] [Medline: [34428216](https://pubmed.ncbi.nlm.nih.gov/34428216/)]
18. The New York Times. Covid-19 Vaccinations: County and State Tracker. 2022. URL: <https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html> [accessed 2022-03-28]
19. Muessig K, Golinkoff J, Hightow-Weidman L, Rochelle A, Mulawa M, Hirshfield S, et al. Increasing HIV testing and viral suppression via stigma reduction in a social networking mobile health intervention among black and Latinx young men and transgender women who have sex with men (HealthMpowerment): protocol for a randomized controlled trial. *JMIR Res Protoc* 2020 Dec 16;9(12):e24043 [FREE Full text] [doi: [10.2196/24043](https://doi.org/10.2196/24043)] [Medline: [33325838](https://pubmed.ncbi.nlm.nih.gov/33325838/)]
20. Teitcher JEF, Bocking WO, Bauermeister JA, Hoefler 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](https://doi.org/10.1111/jlme.12200)] [Medline: [25846043](https://pubmed.ncbi.nlm.nih.gov/25846043/)]
21. Hightow-Weidman L, Muessig K, Claude K, Roberts J, Zlotorzynska M, Sanchez T. Maximizing digital interventions for youth in the midst of Covid-19: lessons from the adolescent trials network for HIV interventions. *AIDS Behav* 2020 Aug;24(8):2239-2243 [FREE Full text] [doi: [10.1007/s10461-020-02870-w](https://doi.org/10.1007/s10461-020-02870-w)] [Medline: [32306214](https://pubmed.ncbi.nlm.nih.gov/32306214/)]
22. Zimmer C, Corum J, Wee S, Kristoffersen M. Covid-19 Vaccine Tracker: Latest Updates. 2022. URL: <https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html> [accessed 2022-08-18]
23. Alvidrez J, Castille D, Laude-Sharp M, Rosario A, Tabor D. The National Institute on Minority Health and Health Disparities research framework. *Am J Public Health* 2019 Jan;109(S1):S16-S20. [doi: [10.2105/ajph.2018.304883](https://doi.org/10.2105/ajph.2018.304883)]
24. Hightow-Weidman L, Budhwani H. A Multidimensional Digital Approach to Address Vaccine Hesitancy and Increase COVID-19 Vaccine Update among African American Young Adults in the South (R01MD016834). 2022. URL: [https://www.nimhd.nih.gov/docs/research\\_framework/nimhdResearchFrameworkAdapt\\_HightowWeidman-Budhwani.pdf](https://www.nimhd.nih.gov/docs/research_framework/nimhdResearchFrameworkAdapt_HightowWeidman-Budhwani.pdf) [accessed 2022-08-18]
25. Shapiro GK, Holding A, Perez S, Amsel R, Rosberger Z. Validation of the vaccine conspiracy beliefs scale. *Papillomavirus Res* 2016 Dec;2:167-172 [FREE Full text] [doi: [10.1016/j.pvr.2016.09.001](https://doi.org/10.1016/j.pvr.2016.09.001)] [Medline: [29074176](https://pubmed.ncbi.nlm.nih.gov/29074176/)]
26. Centers for Disease Control and Prevention (CDC), National Center for Immunization and Respiratory Diseases (NCIRD). Vaccine Confidence Survey Question Bank. 2020. URL: [https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/rca-guide/downloads/CDC\\_RCA\\_Guide\\_2021\\_Tools\\_AppendixD\\_Surveys-508.pdf](https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/rca-guide/downloads/CDC_RCA_Guide_2021_Tools_AppendixD_Surveys-508.pdf) [accessed 2022-03-01]
27. National Institutes of Health (NIH). NIH RADx-UP Common Data Elements - RADx-UP. 2020. URL: <https://radx-up.org/learning-resources/cdes/> [accessed 2022-03-01]
28. Machingaidze S, Wiysonge CS. Understanding COVID-19 vaccine hesitancy. *Nat Med* 2021 Aug 16;27(8):1338-1339. [doi: [10.1038/s41591-021-01459-7](https://doi.org/10.1038/s41591-021-01459-7)] [Medline: [34272500](https://pubmed.ncbi.nlm.nih.gov/34272500/)]
29. Purnell M, Maxwell T, Hill S, Patel R, Trower J, Wangui L, et al. Exploring COVID-19 vaccine hesitancy at a rural historically black college and university. *J Am Pharm Assoc (2003)* 2022;62(1):340-344 [FREE Full text] [doi: [10.1016/j.japh.2021.09.008](https://doi.org/10.1016/j.japh.2021.09.008)] [Medline: [34688566](https://pubmed.ncbi.nlm.nih.gov/34688566/)]
30. Brewer N, Chapman G, Rothman A, Leask J, Kempe A. Increasing vaccination: putting psychological science into action. *Psychol Sci Public Interest* 2017 Dec;18(3):149-207. [doi: [10.1177/1529100618760521](https://doi.org/10.1177/1529100618760521)] [Medline: [29611455](https://pubmed.ncbi.nlm.nih.gov/29611455/)]
31. Brewer NT. What works to increase vaccination uptake. *Acad Pediatr* 2021 May;21(4S):S9-S16. [doi: [10.1016/j.acap.2021.01.017](https://doi.org/10.1016/j.acap.2021.01.017)] [Medline: [33958099](https://pubmed.ncbi.nlm.nih.gov/33958099/)]
32. Brewer NT, Bутtenheim AM, Clinton CV, Mello MM, Benjamin RM, Callaghan T, et al. Incentives for COVID-19 vaccination. *Lancet Reg Health Am* 2022 Apr;8:100205 [FREE Full text] [doi: [10.1016/j.lana.2022.100205](https://doi.org/10.1016/j.lana.2022.100205)] [Medline: [35229080](https://pubmed.ncbi.nlm.nih.gov/35229080/)]
33. Dada D, Djioetio J, McFadden S, Demeke J, Vlahov D, Wilton L, et al. Strategies that promote equity in COVID-19 vaccine uptake for black communities: a review. *J Urban Health* 2022 Feb;99(1):15-27 [FREE Full text] [doi: [10.1007/s11524-021-00594-3](https://doi.org/10.1007/s11524-021-00594-3)] [Medline: [35018612](https://pubmed.ncbi.nlm.nih.gov/35018612/)]
34. Balasuriya L, Santilli A, Morone J, Ainooson J, Roy B, Njoku A, et al. COVID-19 vaccine acceptance and access among black and Latinx communities. *JAMA Netw Open* 2021 Oct 01;4(10):e2128575 [FREE Full text] [doi: [10.1001/jamanetworkopen.2021.28575](https://doi.org/10.1001/jamanetworkopen.2021.28575)] [Medline: [34643719](https://pubmed.ncbi.nlm.nih.gov/34643719/)]

35. Quinn SC, Andrasik MP. Addressing vaccine hesitancy in BIPOC communities — toward trustworthiness, partnership, and reciprocity. *N Engl J Med* 2021 Jul 08;385(2):97-100. [doi: [10.1056/nejmp2103104](https://doi.org/10.1056/nejmp2103104)]
36. Kaiser Family Foundation (KFF). Latest Data on COVID-19 Vaccinations by Race/Ethnicity. 2022. URL: <https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-on-covid-19-vaccinations-by-race-ethnicity/> [accessed 2022-02-21]
37. Centers for Disease Control and Prevention (CDC). CDC COVID Data Tracker. 2021. URL: [https://covid.cdc.gov/covid-data-tracker/#cases\\_casesper100klast7days](https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days) [accessed 2022-08-18]

## Abbreviations

**BYA:** Black or African American young adults

**CDC:** Centers for Disease Control and Prevention

**FDA:** Food and Drug Administration

**NIMHD:** National Institute on Minority Health and Health Disparities

**SCT:** social cognitive theory

*Edited by A Mavragani; submitted 02.05.22; peer-reviewed by K LeMasters, B Eshrati; comments to author 12.07.22; revised version received 04.08.22; accepted 15.08.22; published 02.09.22.*

*Please cite as:*

*Stoner MCD, Browne EN, Tweedy D, Pettifor AE, Maragh-Bass AC, Toval C, Tolley EE, Comello MLG, Muessig KE, Budhwani H, Hightow-Weidman LB*

*Exploring Motivations for COVID-19 Vaccination Among Black Young Adults in 3 Southern US States: Cross-sectional Study*  
*JMIR Form Res* 2022;6(9):e39144

URL: <https://formative.jmir.org/2022/9/e39144>

doi: [10.2196/39144](https://doi.org/10.2196/39144)

PMID: [35969516](https://pubmed.ncbi.nlm.nih.gov/35969516/)

©Marie CD Stoner, Erica N Browne, David Tweedy, Audrey E Pettifor, Allysha C Maragh-Bass, Christina Toval, Elizabeth E Tolley, Maria Leonora G Comello, Kathryn E Muessig, Henna Budhwani, Lisa B Hightow-Weidman. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 02.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# The Effect of Persuasive Design on the Adoption of Exposure Notification Apps: Quantitative Study Based on COVID Alert

Kiemute Oyibo<sup>1,2</sup>, BEng, MSc, PhD; Plinio Pelegrini Morita<sup>1,3,4,5</sup>, PEng, MSc, PhD

<sup>1</sup>School of Public Health Sciences, Faculty of Health, University of Waterloo, Waterloo, ON, Canada

<sup>2</sup>Department of Electrical Engineering and Computer Science, York University, Toronto, ON, Canada

<sup>3</sup>Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

<sup>4</sup>Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

<sup>5</sup>eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

**Corresponding Author:**

Plinio Pelegrini Morita, PEng, MSc, PhD

School of Public Health Sciences

Faculty of Health

University of Waterloo

200 University Avenue West

Waterloo, ON, N2L 3G1

Canada

Phone: 1 5198884567 ext 41372

Email: [plinio.morita@uwaterloo.ca](mailto:plinio.morita@uwaterloo.ca)

## Abstract

**Background:** The adoption of contact tracing apps worldwide has been low. Although considerable research has been conducted on technology acceptance, little has been done to show the benefit of incorporating persuasive principles.

**Objective:** This research aimed to investigate the effect of persuasive features in the COVID Alert app, created by Health Canada, by focusing on the no-exposure status, exposure status, and diagnosis report interfaces.

**Methods:** We conducted a study among 181 Canadian residents, including 65 adopters and 116 nonadopters. This study was based on screenshots of the 3 interfaces, of which each comprised a persuasive design and a control design. The persuasive versions of the first two interfaces supported self-monitoring (of exposure levels), and that of the third interface supported social learning (about how many other users have reported their diagnosis). The 6 screenshots were randomly assigned to 6 groups of participants to provide feedback on perceived persuasiveness and adoption willingness.

**Results:** A multivariate repeated-measure ANOVA showed that there is an interaction among interface, app design, and adoption status regarding the perceived persuasiveness of the interfaces. This resulted in a 2-way ANOVA for each interface. For the no-exposure interface, there was an interaction between adoption status and app design. Among adopters, there was no significant difference  $P=.31$  between the persuasive design (mean 5.36, SD 1.63) and the control design (mean 5.87, SD 1.20). However, among nonadopters, there was an effect of app design ( $P<.001$ ), with participants being more motivated by the persuasive design (mean 5.37, SD 1.30) than by the control design (mean 4.57, SD 1.19). For the exposure interface, adoption status had a main effect ( $P<.001$ ), with adopters (mean 5.91, SD 1.01) being more motivated by the designs than nonadopters (mean 4.96, SD 1.43). For the diagnosis report interface, there was an interaction between adoption status and app design. Among nonadopters, there was no significant difference  $P=.99$  between the persuasive design (mean 4.61, SD 1.84) and the control design (mean 4.77, SD 1.21). However, among adopters, there was an effect of app design ( $P=.006$ ), with participants being more likely to report their diagnosis using the persuasive design (mean 6.00, SD 0.97) than using the control design (mean 5.03, SD 1.22). Finally, with regard to willingness to download the app, pairwise comparisons showed that nonadopters were more likely to adopt the app after viewing the persuasive version of the no-exposure interface (13/21, 62% said yes) and the diagnosis report interface (12/17, 71% said yes) than after viewing the control versions (3/17, 18% and 7/16, 44%, respectively, said yes).

**Conclusions:** Exposure notification apps are more likely to be effective if equipped with persuasive features. Incorporating self-monitoring into the no-exposure status interface and social learning into the diagnosis report interface can increase adoption by >30%.

(JMIR Form Res 2022;6(9):e34212) doi:[10.2196/34212](https://doi.org/10.2196/34212)

**KEYWORDS**

contact tracing app; exposure notification app; COVID Alert; COVID-19; persuasive technology; behavior change; exposure; behavior; effect; design; adoption; use; case study; effectiveness; user interface; mobile phone

## Introduction

### Background

The COVID-19 pandemic resulted in the imposition of public health restrictions and the shutting down of several economies by most national governments worldwide. This necessitated the rollout of digital contact tracing apps to curb the spread of the coronavirus. Digital contact tracing apps help notify users who may have come in contact with someone with COVID-19 so that appropriate safety measures such as self-isolation and testing for COVID-19 can be taken [1]. They were mostly rolled out in high-income countries to support manual methods of contact tracing, which are often labor-intensive, time-consuming, and less likely to be accurate because of the limitation of human memories in recalling contacts [2]. They have the potential to reach a critical mass of adopters and are hence more likely to be effective than traditional means of contact tracing. The emergence of new variants of COVID-19 such as Delta variant [3], which may be resistant to vaccines [4], and its endemic potential are an indication that contact tracing apps may continue to be relevant in the fight against COVID-19 in the long term [5,6]. However, their adoption has been very low and slow owing to several factors [7].

Apart from trust- and privacy-related concerns, the minimalist design of contact tracing apps currently on the Google and Apple app stores tends to limit their perceived usefulness [8]. As noted by Kukuk [9], “[a]part from providing receiving notifications about possible infections, current contract tracing apps appear to not provide a clear benefit to the user.” Digital health experts have identified the lack of persuasive design and motivational affordances as being partly responsible for the low acceptance of contact tracing apps worldwide [7,10]. Research has shown that 56% of the population (eg, in a given country) may have to use contact tracing apps to considerably slow the spread of the virus [11]. Hence, there is a need for researchers to investigate ways to improve the design of contact tracing apps and increase their effectiveness. The minimalist design of contact tracing apps [8,12] (eg, users not being able to track the number of contacts and exposure time) might have been occasioned by the need to minimize collected user data to reduce privacy concerns [13,14] and eliminate fear of government surveillance [15]. Although, this can be seen as an advantage, it has also reduced the usefulness of contact tracing apps [9]. Research has shown that some users may be willing to provide more of their data to contact tracing apps (eg, location data) to receive additional benefits, such as the ability to track the number of daily contacts they had and COVID-19 hot spots [16,17]. The willingness of some users to provide more user data than others to have access to more useful features is an indication of the need for contact tracing apps tailored to different target groups [10,18].

### Persuasive Design

We argued that the incorporation of persuasive features such as self-monitoring, social learning, tailoring, personalization, expertise, praise, and reward has the potential to improve the perceived persuasiveness of contact tracing apps and the reporting of COVID-19 diagnoses [18]. However, there is limited research on the effectiveness of the persuasive design of contact tracing apps in motivating behavior change. Most prior studies [19-21] did not focus on incorporating persuasive features in contact tracing apps. Rather, they focused on the Technology Acceptance Model (TAM), which does not consider persuasive design attributes. From the viewpoint of the TAM, we argue that the perceived usefulness of existing contact tracing and exposure notification apps through persuasive design has been relegated to the background [9,10]. One plausible explanation for this oversight was the need to roll out contact tracing apps as soon as possible to help flatten the curve.

To bridge the gaps in the extant literature, we proposed design guidelines for incorporating persuasive features in exposure notification apps (see our conceptual paper [18]). The guidelines were drawn from the persuasive system design (PSD) model by Oinas-Kekkonen and Harjuma [22], which is commonly used in designing, implementing, and evaluating persuasive systems [23,24]. In this study, we implemented and evaluated the perceived persuasiveness of 2 of the proposed persuasive features (self-monitoring and social learning) from our conceptual paper [18], using the Government of Canada’s COVID Alert app as proof of concept [25]. The app was created by Health Canada in collaboration with Blackberry that provided privacy and security guidance [26]. We chose only 2 persuasive strategies because we could not implement and evaluate all persuasive strategies in the PSD model at the same time, and we had to start from somewhere. In particular, we chose self-monitoring because prior work, such as that by Cruz et al [17], reported that contact tracing app users would like to know the number of persons they have come in contact with. Second, we chose social learning because we believed that learning about the number of other users in your community who have reported their COVID-19 diagnosis holds the potential to motivate users to report theirs when they test positive. Moreover, prior research on persuasive technology has demonstrated that social learning has the capacity to motivate people to engage in beneficial behaviors regardless of culture, gender, or age [27,28]. The rationale for choosing self-monitoring and social learning is discussed in further detail in our prior conceptual paper, which focused on designing exposure notification apps as persuasive technologies [18].

### Study Description

We conducted a survey on Amazon Mechanical Turk among 204 participants residing in Canada to investigate the effect of persuasive design on the adoption and perceived persuasiveness of COVID Alert. This study was based on 2 sets of app designs (persuasive and control), 3 types of use cases (no-exposure

status interface, exposure status interface, and diagnosis report interface), and 2 types of participants (COVID Alert adopters and nonadopters). The persuasive design supports persuasive features, such as self-monitoring and social learning, whereas the control design does not support any persuasive features. Self-monitoring, which is incorporated into the no-exposure and exposure status interfaces of the COVID Alert app, is one of the most commonly used and effective persuasive strategies in behavior change [29-31]. It provides users with opportunities for self-reflection and self-regulation, which result in increased focus and commitment to achieving a target behavior such as social distancing. Moreover, social learning, which is integrated into the diagnosis report interface, is an effective persuasive strategy for motivating behavior change through social influence and pressure [32]. To evaluate the effectiveness of persuasive design, we carried out a 4-factor multivariate repeated-measure ANOVA (RM-ANOVA) [33] based on interface, app design, adoption status, and perceived persuasiveness. Our overall hypothesis is that the persuasive design of exposure notification apps, regardless of the use case (interface), is more likely to be persuasive and adopted by potential users than the control design. Moreover, we hypothesize that adopters are more likely to find exposure notification apps persuasive than nonadopters, regardless of app design and use case.

## Related Work

### Overview

Before conducting this research, we searched 6 databases (Scopus, CINAHL, PubMed [MEDLINE], IEEE Xplore Digital Library, ACM Digital Library, and Web of Science) between October 30, 2020, and November 20, 2020, using the following terms: (*contact tracing* OR *contact-tracing* OR *exposure notification* OR *exposure-notification* OR *contact notification* OR *contact-notification* OR *GAEN*) AND (*app* OR *apps* OR *application\** OR *technolog\** OR *system* OR *systems*) AND (*percept\** OR *adopt\** OR *accept\** OR *uptake* OR *use* OR *usage*) AND (*covid\** OR *coronavirus* OR *SARS-CoV-2*). In addition, we searched Google Scholar between November 21, 2020, and January 31, 2021, using terms such as *COVID-19 contact tracing app* and *COVID-19 exposure notification app*. The systematic review, which uncovered the key factors that drive the acceptance of contact tracing apps, is published in *Frontiers in Digital Health* [34]. The protocol for this review was published in the *Journal of Medical Internet Research* [35]. In this study, we review the key related articles retrieved from the database search, focusing on privacy, trust, and persuasive design.

### Privacy and Trust

Privacy and trust are among the top-ranking ethical issues that COVID-19 stakeholders such as researchers, designers, and the public are concerned with when it comes to digital contact tracing [36-38]. In the context of web-based systems, privacy refers to the level of protection and security of user data and interaction while using an electronic system connected to the internet. It entails the collection, storage, use, and sharing of a user's personal information [39]. In contrast, trust (despite not having a universally accepted scholarly definition [40]), in the context of web-based activities, is regarded as a cognitive mechanism adopted by users when interacting with

internet-connected systems. Usually connected to the perceived quality, usability, and expertise of a web-based system such as a website, trust "operates to reduce the amount of [perceived] risk by reducing perceptions of anxiety and uncertainty" [40]. Preliminary research shows that there is a significant relationship between privacy concerns and trust, with each having the potential to impact the adoption of web-based systems, such as social networking sites [41,42] and e-commerce sites [43,44]. For example, Zlatolas et al [41] found that the higher the perceived privacy risk of using Facebook, the lower the perceived trust of users, and the lower the perceived trust in a social media site, the higher the privacy concerns of users. Trust is often associated with the success or failure of an e-commerce website, as web-based shoppers are concerned with unsafe products, insecure payment methods, loss of privacy, identity theft, and misuse of personal information [45].

In the contact tracing domain, research has also shown that privacy concerns and trust can impact the adoption of contact tracing apps [38]. For example, Sharma et al [19], Altmann et al [21], Kaspar [46], and Velicia-Martin et al [47] found in their work on technology acceptance that the higher people's concern about privacy is, the less likely they are to download, install, or use contact tracing apps. Moreover, Sharma et al [19], Altmann et al [21], and Kaspar [46] found that the higher the users' perceived trust in contact tracing apps and their stakeholders, such as the government, the higher their likelihood of adopting them. In contrast, Jonker et al [48] and Thomas et al [49] found that the higher the distrust of users (eg, in governments and tech companies [50]), the less likely they are to adopt contact tracing apps. Hence, as a way of enacting privacy protection, Jonker et al [48] recommended that governments implement contact tracing apps with adequate realistic privacy-preserving features; for example, users should be given control over their data, including deciding what data they want to share, whom they want to share it with, how and when they want to share it, and what it will be used for. Similarly, Walrave et al [20] recommended that contact tracing app sponsors inform potential users about the data to be collected and minimize data collection and the amount of time required to read and evaluate privacy terms by using visual presentation to improve comprehension. Finally, in furthering and fostering public trust, Altmann et al [21] recommended that national governments around the world should consider delegating the mandate of digital contact tracing to reputable and transparent public health institutions, over which they have little to no control.

### Persuasive Design

Although a substantial amount of work has been done with regard to the impact of privacy and trust on contact tracing app adoption (as shown in the previous subsection), little has been done with regard to the impact of persuasive design. As of the time of writing this paper, we found only 2 studies [17,48] that investigated the benefit of incorporating persuasive features in contact tracing apps. One of the studies (Cruz et al [17]) found that more than half of the participants wanted to know how many infected people they had come in contact with (including the location and time) by way of self-monitoring. The study also found that most participants were more willing to share

their locations when they were offered tangible rewards [17]. Similarly, another study (Jonker et al [48]) found that participants preferred contact tracing apps that offer tangible rewards, such as money and free COVID-19 testing. However, these studies were primarily based on contact tracing app descriptions and not implementations. Moreover, these studies were not based on a comparative analysis of intervention designs (equipped with persuasive strategies) or control designs (unequipped with persuasive strategies). Most importantly, the studies were carried out in the first half of 2020, when many people were less familiar with or had not used contact tracing and exposure notification apps. Hence, there is a need for this study to bridge the gap in the extant literature regarding the effect of persuasive design on contact tracing and exposure notification app design.

## Methods

In this section, we focus on app design, measurement instruments, recruitment of participants, experimental design and data analysis, sample size calculation, and research model and hypotheses.

### App Design

COVID Alert is the Government of Canada's official app for contact tracing and exposure notification. Released on July 31, 2020, it uses Google/Apple Exposure Notification application programming interfaces to enforce strong privacy measures. Hence, it does not track the user's location or collect personally identifiable information such as name, contacts, address, or health information. Similar to many exposure notification apps on the market, the COVID Alert app (persuasive or control design) comprises 3 key use cases: no-exposure status interface, exposure status interface, and diagnosis report interface (Figures 1 and 2). In the persuasive design, we implemented 2 types of persuasive strategies (self-monitoring and social learning) drawn from the PSD model [22]. The PSD model is a framework for the design, implementation, and evaluation of persuasive systems. It comprises 28 persuasive strategies. In our conceptual paper on exposure notification app design [18], we discuss likely persuasive strategies from the PSD model that can be incorporated into exposure notification apps to make them more effective and appealing. These include self-monitoring, tailoring, social learning, normative influence, trustworthiness, and authority. The rationale for implementing these strategies is described in the conceptual paper. In this study, we implemented the aforementioned strategies by focusing on self-monitoring (incorporated into the no-exposure and exposure status

interfaces) and social learning (incorporated into the diagnosis report interface).

As shown in Figure 1, the no-exposure status interface informs the user that they have not been exposed to COVID-19 by being close to someone with COVID-19 in the last 14 days. The exposure status interface notifies the user that they may have been exposed to COVID-19 by being in close contact with someone with COVID-19, and provides information on what to do next (eg, self-isolate or go test for COVID-19 in the event of having symptoms). Finally, the diagnosis report interface enables a user who has tested positive to enter a one-time key given to them by the public health authority. We regard these 3 key original interfaces of the COVID Alert app, which are not equipped with persuasive features, as control designs (Figure 1).

Figure 2 shows the corresponding persuasive designs equipped with persuasive features. The no-exposure and exposure status interfaces are equipped with self-monitoring, and the diagnosis report interface is equipped with social learning. Self-monitoring is a persuasive feature that allows users to track their COVID-19 exposure levels over time. Figure 3 [34,51,52] illustrates the operational mechanism of self-monitoring. A person observes their own behavior and reflects on it, as though they are looking at themselves in the mirror. If they are not impressed with what they see (in the mirror), they regulate themselves by improving on the target behavior [29,53,54]. In the no-exposure status interface, users can track total and average number of daily contacts and minutes exposed. In the exposure status interface, users can view the cumulative sum of contacts and exposure minutes in the last 14 days within which they must have been exposed. It is hoped that by seeing these summary statistics, users will be motivated to regulate their social distancing behavior. In contrast, social learning is a persuasive feature that allows users to be aware of other people's behavior in the hope that they will be socially pressured and motivated to adopt the observed behavior. Figure 3 illustrates the operational mechanisms of social learning [53,55,56]. Social learning is based on the premise that observational learning cannot occur unless cognitive processes that mediate the learning process occur [52]. Figure 3 demonstrates that by observing others' behavior, one is motivated through social pressure to imitate the observed behavior for the common good. In the diagnosis report interface, the app informs the user about the number of users who have reported their COVID-19 diagnosis on a given day in the hope that they would be socially pressured to report if they tested positive to promote public health safety.

Figure 1. Control designs of the 3 key interfaces of the COVID Alert app.

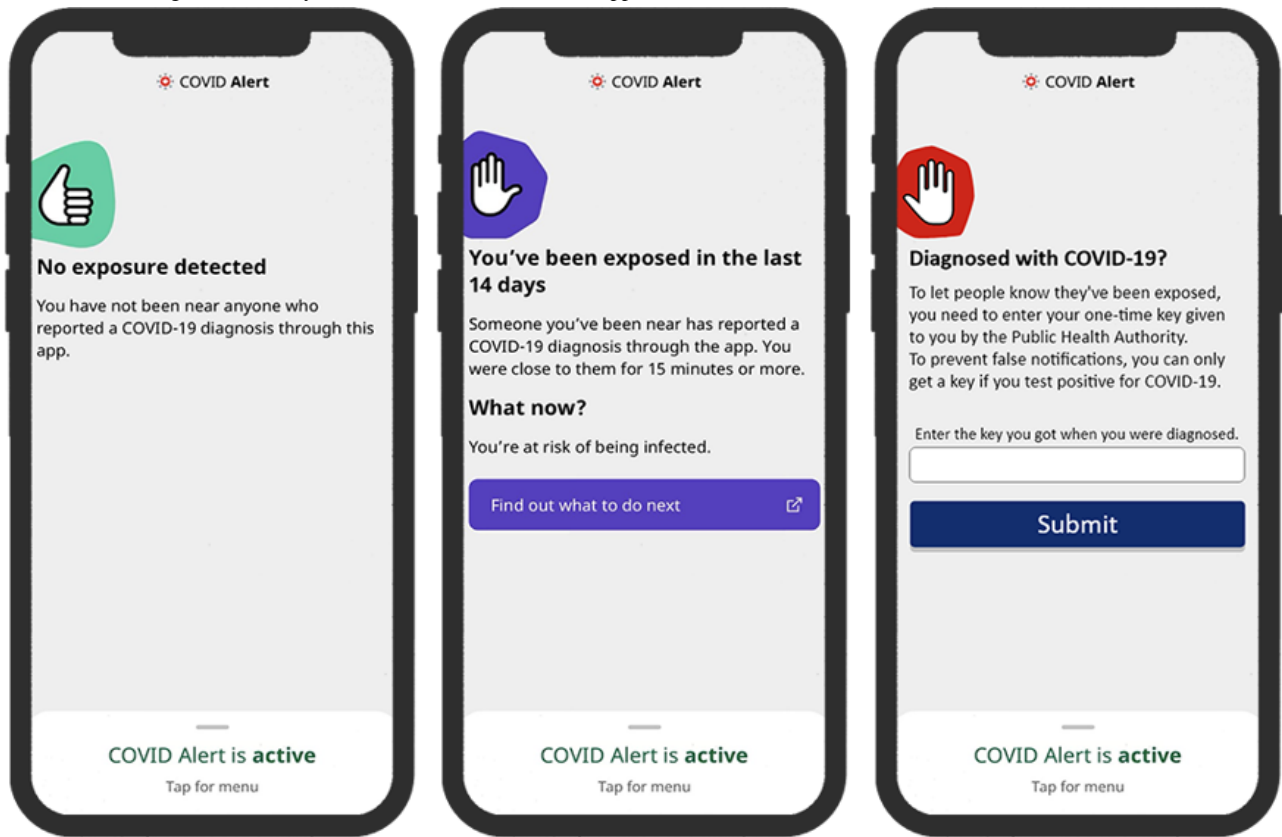
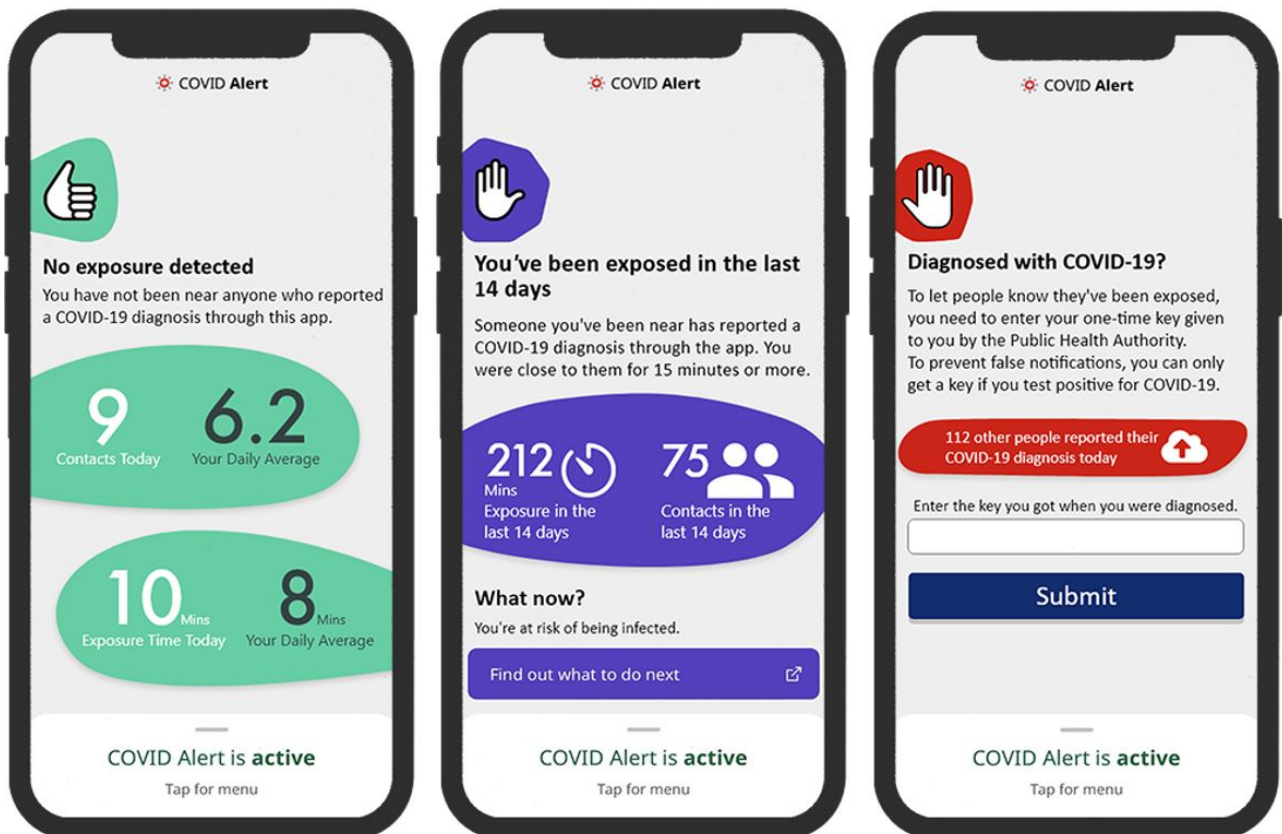
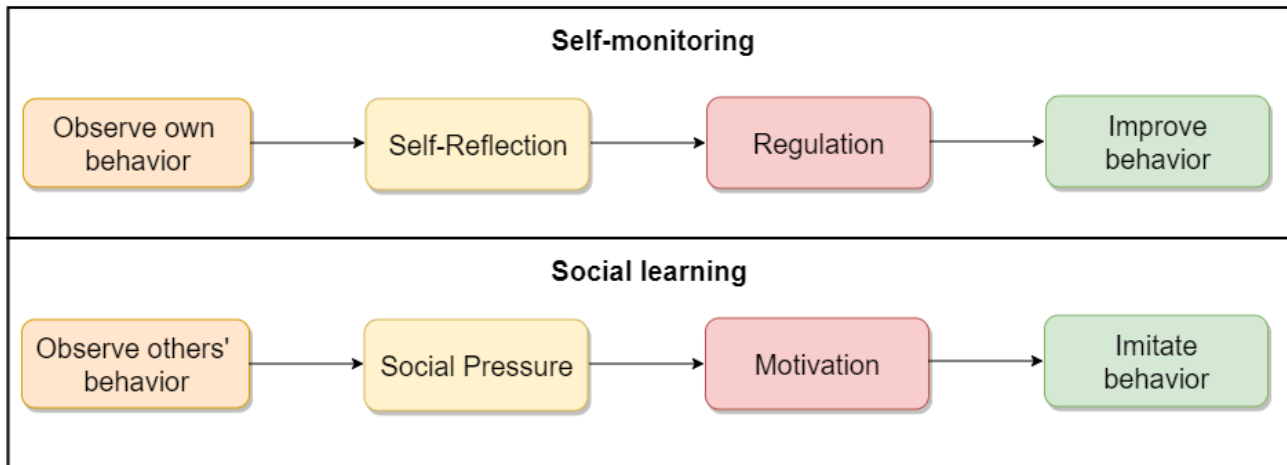


Figure 2. Persuasive designs of the 3 key interfaces of the COVID Alert app.



**Figure 3.** The operational mechanism of self-monitoring and social learning [34,51,52].



**Measurement Instruments**

To investigate the effectiveness of the persuasive design, we measured 2 key constructs of interest: perceived persuasiveness of each of the interfaces (shown in Figures 1 and 2) and participants’ willingness to download the COVID Alert app from the app store. Table 1 shows the measures for both the

constructs. Perceived persuasiveness refers to and measures the ability of the visual and informational design of an app to motivate users to adopt it. In this study, perceived persuasiveness is a reflective measure that captures how well the visual design of the COVID Alert app convinces and influences the user to start or continue using the app.

**Table 1.** Measurement instruments.

Construct	Items measuring construct
Perceived persuasiveness (“strongly disagree: 1” to “strongly agree: 7”) [57]	The app design (name of interface)... 1. ...influences me to start or continue using the COVID Alert app. 2. ...is convincing for me to start or continue using the COVID Alert app. 3. ...is relevant to my using or continued use of the COVID Alert app.
Willingness to download app from store (yes or no)	Now that I know about the COVID Alert app as the Government of Canada’s official exposure notification app, I will download it from the Apple or Google store to slow down the spread of the coronavirus.
Adoption status	Which of the following best describes you? 1. I am currently using the COVID Alert app. 2. I am currently using a COVID-19 contact tracing or exposure notification app other than COVID Alert. 3. I am not currently using any COVID-19 contact tracing or exposure notification app.

In the context of this study, perceived persuasiveness can be viewed as a proxy for the TAM or Theory of Planned Behavior constructs such as perceived usefulness [56,58], perceived compatibility with existing experiences, values, and tasks [59,60], and peer or superior influence [61], which have the potential to impact the adoption of new technologies. For example, the more a new technology is perceived as useful and compatible with the user’s past experiences, values, and tasks, the more relevant they will deem it and the more likely they will be to adopt it [61]. However, although perceived persuasiveness may be associated with constructs such as perceived ease of use and perceived usefulness [57,58], perceived compatibility with tasks [59], and social influence [62], it is not synonymous with any of these constructs. For example, the fact that a user perceives an app to be persuasive (motivating) may not mean that they find it easy to use, useful, or compatible with prior experiences, values, and tasks or vice versa. One plausible explanation is that some users may perceive an app (eg, a game) to be persuasive based on hedonic

characteristics (such as perceived aesthetics [63] and perceived enjoyment [64]), without considering the utilitarian (eg, perceived usefulness) or compatibility features. In contrast, other users may perceive an app (eg, an exposure notification app) to be persuasive based on utilitarian or compatibility features without paying much attention to hedonic features. In the context of the PSD model, perceived persuasiveness can be viewed as a proxy for the four main categories of persuasive strategies. They include primary task support, dialog support, social support, and credibility support, which have direct and indirect relationships with perceived persuasiveness and adoption intention, respectively [65]. In particular, primary task support (defined as persuasive features that enable users to realize the main goal of a persuasive system) can be compared to perceived usefulness in the TAM. For example, in the work by Lehto et al [65], based on a web-based persuasive health system, primary task support was operationalized using utility-oriented items including (1) the system provides me with means to lose weight, (2) the system helps me lose weight, and

(3) the system helps me change my eating habits, which reflect perceived usefulness.

For this study, the perceived persuasiveness measure was adapted from the work by Lehto et al [65], to suit the context of exposure notification apps. It is a 7-point scale ranging from *strongly disagree* (1) to *strongly agree* (7). Moreover, willingness to download refers to and measures participants' intention to adopt the app to curb the spread of the coronavirus after seeing or learning about its functionality. It was based on a *yes-or-no* measure. Finally, we measured adoption status by asking participants to choose 1 of the 3 options shown in [Table 1](#). If they chose the first and third options, they were regarded as COVID Alert adopters and nonadopters, respectively. Those who chose the second option were filtered out of the data analysis, as we were interested in analyzing and comparing participants who had installed and interacted with the COVID Alert app and those who had not in the past.

### Participants

The criterion for inclusion in the study was that participants must be residents of Canada, regardless of sex, gender, age, education, country of origin, and contact tracing app adoption status. We did not place any demographic restrictions on who could participate in the study because everyone, regardless of the enumerated demographic variables, is liable to be exposed to COVID-19, and is thus expected to use exposure notification apps such as COVID Alert. We recruited participants residing in Canada with at least one year of smartphone use experience on Amazon Mechanical Turk to evaluate the persuasive and control designs of the COVID Alert app. Amazon Mechanical Turk is an inexpensive crowdsourcing web-based commercial platform for recruiting a nonconvenience sample of participants worldwide. Research has shown that owing to its

quality-assurance mechanism, the platform has the potential to yield high-quality data [66]. The recruitment of study participants took place between December 25, 2020, and January 25, 2021. With the aid of our laboratory-wide account, the first author used the requester interface to post details of the study on the Amazon Mechanical Turk platform. The requester interface allows the researcher to specify the number of participants, duration of the study, and types of participants using filtering terms such as country and location [67]. We tweaked the default JavaScript code in the requester interface to randomly assign 1 of the 6 exposure notification app interfaces to each potential anonymous participant. Hence, each participant only viewed the interface assigned to them as described in [Multimedia Appendix 1](#), without interacting with it. Before completing the web-based questionnaire, each participant was requested to read the information and consent forms and provide informed consent. Upon consent, participants were allowed to complete the survey; otherwise, they were directed to the end of the survey. Each participant was remunerated with US \$2 in appreciation of their time.

A total of 204 participants took part in the study. Of these, 65 (32%) had already used the COVID Alert app, 17 (8%) were using other contact tracing apps, 116 (57%) did not use the COVID Alert app or any other contact tracing app at the time of taking the survey, and 6 (3%) did not specify their adoption status. The first and third subgroups were regarded as the COVID Alert adopter group (n=65) and the nonadopter group (n=116), respectively. The second and fourth subgroups (n=23) were filtered out during data analysis. [Table 2](#) shows the demographics of the COVID Alert adopters and nonadopters (n=181) assigned to the 6 user interfaces, comprising 3 control designs (C1, C2, and C3) and 3 persuasive designs (P1, P2, and P3).

**Table 2.** Participants' demographics based on the 6 user interfaces (N=181).

Criterion and subgroup	Overall users, n	No-exposure interface, n		Exposure interface, n		Diagnosis report interface, n	
		C1 <sup>a</sup>	P1 <sup>b</sup>	C2	P2	C3	P3
<b>Gender</b>							
Male	106	16	20	18	21	19	12
Female	73	10	12	12	16	9	14
Others	2	1	0	0	0	1	0
<b>Age (years)</b>							
<18	1	0	1	0	0	0	0
18 to 24	36	1	6	7	10	5	7
25 to 34	64	8	10	12	12	11	11
35 to 44	48	9	9	6	9	10	5
45 to 54	19	6	2	3	4	2	2
>55	10	2	3	1	2	1	1
Unspecified	3	1	1	1	0	0	0
<b>Education</b>							
Technical or trade	5	0	0	1	2	1	1
High school	39	2	11	4	9	3	10
Bachelor's	99	20	14	18	19	18	10
Master's	29	3	4	6	6	5	5
Doctorate	3	1	1	0	0	1	0
Other	6	1	2	1	1	1	0
<b>Using smartphone (years)</b>							
1 to 5	27	3	6	2	4	5	7
6 to 10	86	14	16	18	19	9	10
11 to 20	59	8	9	10	12	13	7
>20	8	2	1	0	2	2	1
Unspecified	1	0	0	0	0	0	1
<b>Country of origin</b>							
Canada	143	24	21	24	31	21	22
Other	38	3	11	6	6	8	4
<b>Adoption status</b>							
Adopters	65	10	11	11	11	13	9
Nonadopters	116	17	21	19	26	16	17

<sup>a</sup>C: control design.

<sup>b</sup>P: persuasive design.

## Experimental Design and Data Analysis

This study was based on a web-based questionnaire in which each participant was randomly assigned to 1 of the 6 user interfaces shown in [Figures 1](#) and [2](#). Before questions were asked to the participants, the functionality of the COVID Alert app was described to them (see [Multimedia Appendix 1](#) for details on the experimental design and accompanying information presented to participants). Two types of data analysis were carried out: path modeling and multivariate

RM-ANOVA [33]. First, the path modeling set out to uncover the strength of the relationship between the perceived persuasiveness of each of the 3 interfaces (no-exposure status, exposure status, and diagnosis report) and the willingness to download the app by nonadopters. This analysis helped us establish that there is a significant relationship between the perceived persuasiveness of an exposure notification app and the willingness to adopt it by nonadopters.



Second, the experimental design, based on a 4-way multivariate RM-ANOVA factorial design, aimed to understand the main effect of *app design*, *interface*, and *adoption status* on the *perceived persuasiveness* of each user interface and their interactions. On the basis of this 4-way multivariate RM-ANOVA factorial design, we aimed to understand the main effect of the first 3 variables on the perceived persuasiveness of each of the 3 user interfaces and their interactions. The app design has 2 conditions (*persuasive* and *control*), the interface has 3 levels (*no-exposure status*, *exposure status*, and *diagnosis report*), and the adoption status has 2 levels (*adopters* and *nonadopters*). Moreover, perceived persuasiveness was measured repeatedly using 3 indicators as shown in [Table 1](#). Finally, among the nonadopter group, we investigated the effect of app design on participants' willingness to download the COVID Alert app from the app store. Using 2x2 chi-square tests [68], we compared, for each user interface, the percentage of participants who viewed the persuasive design that said "yes" with the percentage of participants who viewed the control design that said "yes". This pairwise comparison helped to uncover any significant difference between the persuasive and control design groups.

### Sample Size Calculation

Before conducting this study, we computed the sample size using the University of British Columbia's web-based power and sample size calculator developed by Brant [69]. We chose the default significance level of .05 and a power level of 0.80. Moreover, we chose our SD value to be 1.0, and the mean difference between the 2 groups as 0.8 on a 7-point Likert scale (ie, >10% difference). The SD was derived from a similar study of the principles of persuasion by Cialdini, conducted among individualist participants from North America [70]. In particular, the SD for the liking principle, which is highly related to the perceived persuasiveness construct in this study, was 1.09. Hence, we decided to use a SD of approximately 1.0 for the calculation of our sample size for each group. The calculation (based on a 2-sided test) resulted in a sample size of 25 for each group. As shown in [Table 2](#), a total of 6 groups met this sample size requirement, with 5 of them being >30.

### Research Model and Hypotheses

We based our data analysis on path modeling and multivariate RM-ANOVA. [Figure 4](#) shows the hypothesized model. This model was based on prior research, which showed that there is a significantly strong relationship between the perceived persuasiveness of an app (such as a fitness app) and adoption intentions [57]. On the basis of this finding and the fact that screenshots of key interfaces of an app are often included in its description in the app store, we hypothesized as follows: *hypothesis H1: the higher the perceived persuasiveness of an exposure notification app in the app store, the more likely users will download it*. This hypothesis is based on the premise that potential users will be able to view the key interfaces of the app (in addition to reading its description) in the app store before making their decision to download it. It is broken down for each of the 3 key user interfaces as follows:

1. H1a: the higher the perceived persuasiveness of the no-exposure status interface in the app store, the more likely users will download the COVID Alert app.
2. H1b: the higher the perceived persuasiveness of the exposure status interface in the app store, the more likely users will download the COVID Alert app.
3. H1c: the higher the perceived persuasiveness of the diagnosis report interface in the app store, the more likely users will download the COVID Alert app.

In addition, using an exploratory approach, we investigated which of the 3 interfaces (ie, perceived persuasiveness) has the strongest effect on users' willingness to download the COVID Alert app. It is noteworthy that we do not imply or mean a causal-effect relationship in H1 or each time we use the word *effect* in characterizing the relationship between perceived persuasiveness and willingness to download the app. As the mantra goes, *correlation does not mean causation*. Moreover, we hypothesized that the perceived persuasiveness of each interface will be influenced by the app design. In other words, given that persuasive designs support persuasive features such as self-monitoring and social learning, we hypothesized as follows:

1. H2a: the perceived persuasiveness of the persuasive design of the no-exposure status interface will be higher than that of the control design.
2. H2b: the perceived persuasiveness of the persuasive design of the exposure status interface will be higher than that of the control design.
3. H2c: the perceived persuasiveness of the persuasive design of the diagnosis report interface will be higher than that of the control design.

Third, research shows that adopters perceive and rate new technologies more favorably than nonadopters [71-73]. For example, Dickerson and Gentry [73] found that prior experience with other computer-related products and services played a significant role in the movement of people toward the purchase of a home computer. Hence, we hypothesized that the perceived persuasiveness of each interface will be influenced by app adoption status. In other words, given that users of COVID Alert (adopters) are familiar with and are currently using it to track their exposure, they are more likely to evaluate it favorably. Hence, we hypothesized as follows:

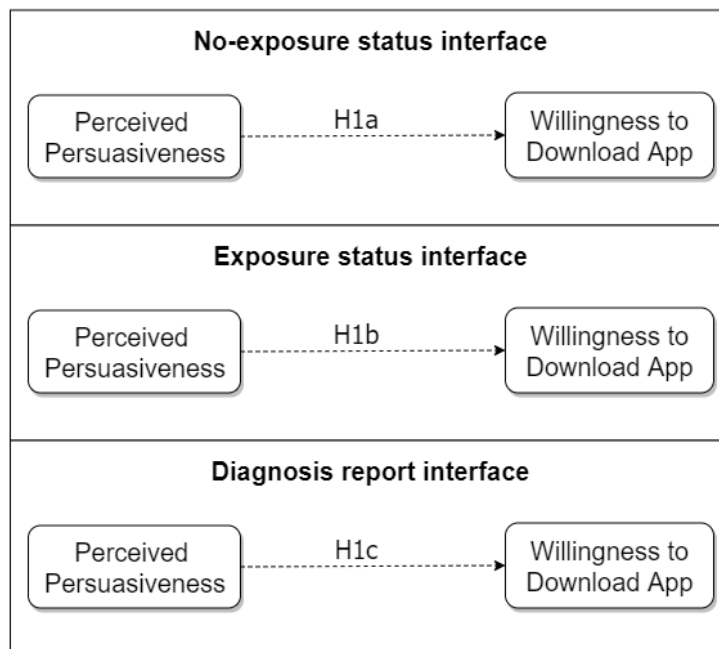
1. H3a: adopters are more likely to perceive the no-exposure status interface to be persuasive than nonadopters.
2. H3b: adopters are more likely to perceive the exposure status interface to be persuasive than nonadopters.
3. H3c: adopters are more likely to perceive the diagnosis report interface to be persuasive than nonadopters.

Fourth, given the hypothesized relationship between perceived persuasiveness and willingness to download the app (H1), we hypothesized that persuasive versions are more likely to be downloaded by nonadopters than control versions (H4). Some nonadopters, before the completion of the study, might have refused to download the control version of the COVID Alert app in the past for various reasons. However, with the integration of persuasive features such as self-monitoring and social learning, which provide some utilitarian benefit

(monitoring of exposure levels) and a socially motivational message, we hypothesized as follows:

1. H4a: nonadopters who viewed the persuasive design of the no-exposure status interface are more likely to adopt the COVID Alert app than those who viewed the control design.
2. H4b: nonadopters who viewed the persuasive design of the exposure status interface are more likely to adopt the COVID Alert app than those who viewed the control design.
3. H4c: nonadopters who viewed the persuasive design of the diagnosis report interface are more likely to adopt the COVID Alert app than those who viewed the control design.

**Figure 4.** Research model for the relationship between perceived persuasiveness and willingness to download the app by nonadopters. H: hypothesis.



### Ethics Approval

This study was approved by the University of Waterloo Research Ethics Committee (ORE 42638).

### Results

In this section, we present the results based on our hypotheses. The results include the data-driven model, the mean values of perceived persuasiveness for each of the 3 interfaces, the ANOVA to uncover the main effects and interactions of factors, and the percentages of nonadopters who are willing to download the COVID Alert app from the Apple or Google store because of their awareness of it through the survey.

### Data-Driven Path Model

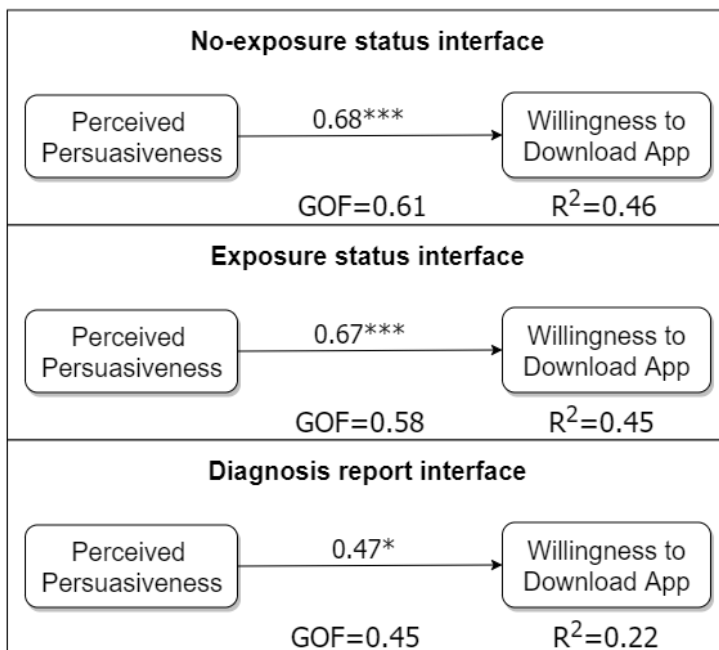
Figure 5 shows the data-driven models for the 3 key user interfaces. The models aim to answer the first set of hypotheses (H1a to H1c). They were built using the *partial least-squares* path modeling package in RStudio [74]. The no-exposure status interface model was built using a subset of the C1 and P1 participants (n=38) who were nonadopters, as shown in Table 2. The other 21 participants did not respond to the question on willingness to download the app. Similarly, the exposure status interface model was built using only the C2 and P2 nonadopters (n=45). Finally, the diagnosis report interface model was built using only the C3 and P3 participants (n=33). As shown in Table 1, one item was used to measure the willingness to download the app, and 3 items were used to measure perceived persuasiveness. In constructing the models, the responses *yes* and *no* to willingness to download the app were coded as 1 and

0, respectively. All the construct items were treated as reflective indicators in the measurement models. Unlike formative indicators, which are considered the causes or drivers of the construct (ie, latent variable) that they measure, reflective indicators are considered to be caused by the construct that they measure [75]. Before analyzing the structural models, we evaluated the measurement models to ensure that the required preconditions such as indicator reliability, internal consistency reliability, convergent validity, and discriminant validity of the multiitem construct are satisfied. The outer loading metric was used to measure indicator reliability, which was >0.7 for most of the indicators that measured perceived persuasiveness in the 3 models. However, in the second model, the third indicator (*The app design is relevant to my using or continued use of the COVID Alert app*) had an outer loading value of 0.64. In the third model, the indicator was removed because its outer loading value was <0.40. The Dillion-Goldstein metric was used to assess the internal consistency reliability of perceived persuasiveness, which was also >0.7. The average variance extracted metric was used to assess the convergent validity of perceived persuasiveness, which was >0.5. Finally, the cross-loading metric was used to assess the discriminant validity of perceived persuasiveness. Its indicators loaded higher on itself than on willingness to download the app [74].

Overall, regardless of the interface, the relationship between perceived persuasiveness and willingness to download an app was statistically significant with  $\beta > 0.40$ . We also conducted a multigroup analysis to determine the significant difference between each pair of path coefficients in the 3 submodels. The results showed no significant difference between each pair,

although the path coefficients for the no-exposure status ( $\beta=.67$ ;  $P<.001$ ) were numerically higher than those of the interface ( $\beta=.68$ ;  $P<.001$ ) and the exposure status interface diagnosis report interface ( $\beta=.47$ ;  $P=.04$ ).

**Figure 5.** Data-driven model based on each of the 3 key user interfaces. GOF: goodness of fit. \* $P<.05$ ; \*\*\* $P<.001$ .



### Mean Values of Perceived Persuasiveness and RM-ANOVA

#### Overview

In this section, we address the second and third sets of hypotheses (ie, H2 and H3) by conducting a 4-factor multivariate

RM-ANOVA based on the interface, app design, adoption status, and perceived persuasiveness. The results of the analysis (Table 3) show a main effect of adoption status ( $F_{507,1}=28.94$ ;  $P<.001$ ) and an interaction between interface, adoption status, and app design ( $F_{507,2}=5.90$ ;  $P=.002$ ). Owing to the interaction, we carried out a 2-way ANOVA taking each interface, app design, and adoption status at a time.

**Table 3.** Repeated-measure ANOVA based on interface, adoption status, app design, and perceived persuasiveness.

	Adoption status	Interface×adoption status×app design
Df Res <sup>a</sup>	507	507
F (df)	28.94 (1)	5.90 (2)
P value	<.001	.002

<sup>a</sup>Df Res: degree of freedom residual.

#### Two-Way ANOVA for Each Interface

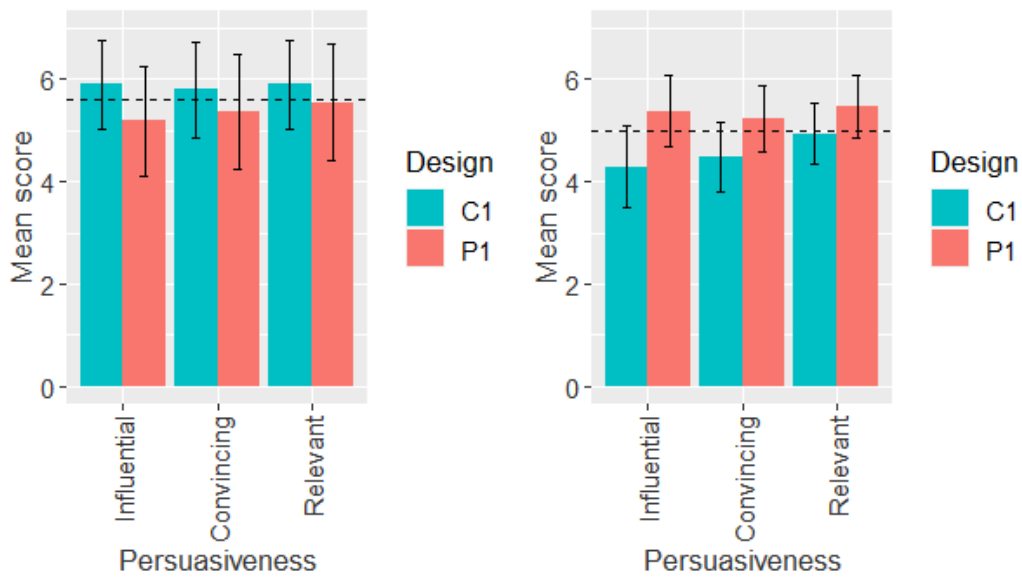
In this section, owing to the 3-way interaction shown in Table 3, we conducted a 2-way ANOVA based on the adoption status and app design for each of the 3 interfaces.

##### No-Exposure Status Interface

Figure 6 shows the mean ratings of perceived persuasiveness of the no-exposure status interface for adopters and nonadopters. Overall, adopters rated the interface higher than nonadopters. As shown in Table 4, the 2-way ANOVA showed that there was a main effect of adoption status ( $F_{173,1}=10.82$ ;  $P=.001$ ) and an interaction between adoption status and app design ( $F_{173,1}=6.93$ ;  $P=.009$ ).

Owing to the interaction between adoption status and app design, we carried out a further 1-way ANOVA at each level of adoption status and app design as shown in Table 5. The results showed that there was a main effect of app design within the nonadopter group, with a medium effect size ( $F_{112,1}=12.34$ ;  $P<.001$ ;  $\eta_p^2=0.10$ ). The persuasive design (mean 5.37, SD 1.30) had a significantly higher mean value than the control design (mean 4.57, SD 1.19) did. Moreover, adoption status had a main effect regarding the control design, with a large effect size ( $F_{79,1}=20.41$ ;  $P<.001$ ;  $\eta_p^2=0.21$ ). The adopter group (mean 5.87, SD 1.20) rated the control design significantly higher than the nonadopter group (mean 4.57, SD 1.19).

**Figure 6.** Mean ratings of perceived persuasiveness of the no-exposure interface for the COVID Alert adopters and nonadopters. Horizontal bar represents overall mean value of perceived persuasiveness. Vertical bars represent 95% CIs. C: control design; P: persuasive design.



**Table 4.** Two-way ANOVA based on adoption status and app design for the no-exposure status interface.

	Adoption status	Adoption status×app design
<i>Df Res</i> <sup>a</sup>	173	173
<i>F (df)</i>	10.82 (1)	6.93 (1)
<i>P value</i>	.001	.009

<sup>a</sup>*Df Res*: degree of freedom residual.

**Table 5.** Further 1-way ANOVA for the perceived persuasiveness of the no-exposure status interface at each level of adoption status and app design (small effect size:  $\eta_p^2=0.01$ ; medium effect size:  $\eta_p^2=0.06$ ; larger effect size:  $\eta_p^2=0.14$ ) [76].

	One-way ANOVA for each app design		App design effect
	C1 <sup>a</sup>	P1 <sup>b</sup>	
<b>One-way ANOVA within each adoption status</b>			
Adopter	5.87	5.36	$F_{61,1}=1.05$ ; $P=.31$
Nonadopter	4.57	5.37	$F_{112,1}=12.34$ ; $P<.001$ ; $\eta_p^2=0.10$
Adoption effect	$F_{79,1}=20.41$ ; $P<.001$ ; $\eta_p^2=0.21$	$F_{94,1}=0.04$ ; $P=.84$	N/A <sup>c</sup>

<sup>a</sup>C: control design.

<sup>b</sup>P: persuasive design.

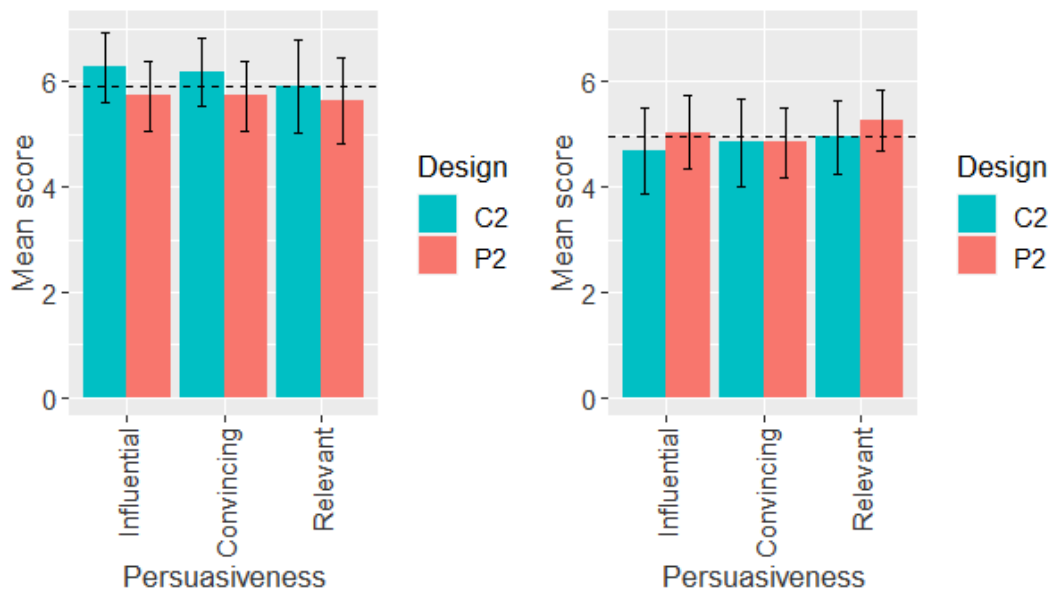
<sup>c</sup>N/A: not applicable.

**Exposure Status Interface**

Figure 7 shows the mean rating of the perceived persuasiveness of the exposure status interface for adopters and nonadopters. The 2-way ANOVA based on adoption status and app design

(Table 6) showed that there was only a main effect of adoption status ( $F_{197,1}=19.03$ ;  $P<.001$ ) with a medium effect size ( $\eta_p^2=0.09$ ). In other words, the adopters significantly rated the perceived persuasiveness of the interface (mean 5.91, SD 1.01) higher than the nonadopters (mean 4.96, SD 1.43).

**Figure 7.** Mean scores of perceived persuasiveness of the exposure status interface for COVID Alert adopters and nonadopters. Horizontal bar represents the overall mean value of the construct for each user group. Vertical bars represent 95% CIs. C: control design; P: persuasive design.



**Table 6.** Two-way ANOVA based on app design and adoption status for the exposure status interface (small effect size:  $\eta^2=0.01$ ; medium effect size:  $\eta^2=0.06$ ; larger effect size:  $\eta^2=0.14$ ) [76].

	Adoption status	Adoption status×app design	App design
<i>Df Res</i> <sup>a</sup>	197	197	197
<i>F (df)</i>	19.03 (1)	1.81 (1)	0.40 (1)
<i>P value</i>	<.001	0.18	0.53

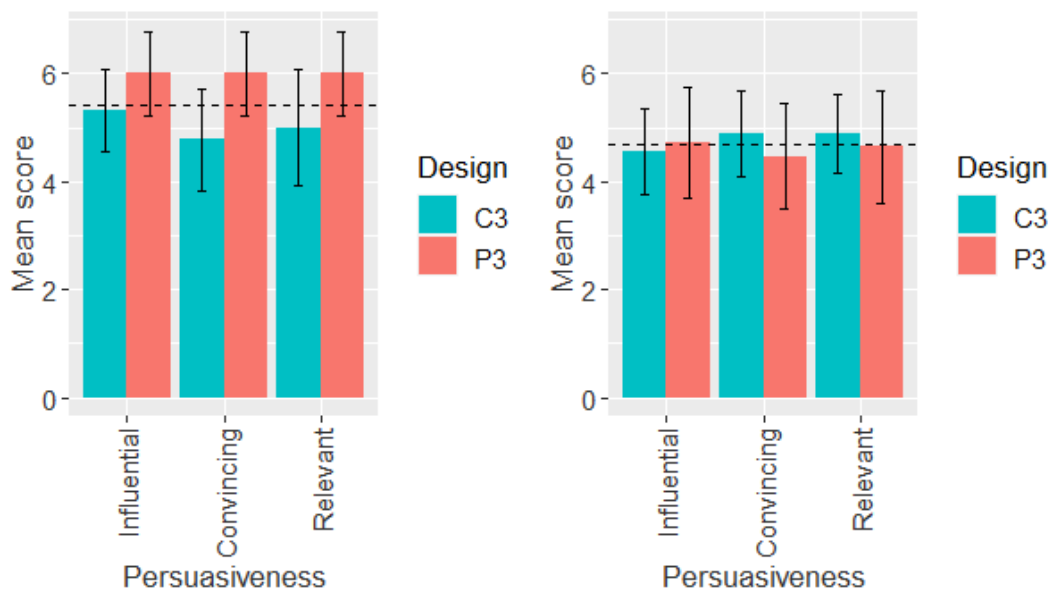
<sup>a</sup>*Df Res*: degree of freedom residual.

**Diagnosis Report Interface**

Figure 8 shows the mean rating of the perceived persuasiveness of the diagnosis report interface for the adopter and nonadopter

groups. The 2-way ANOVA based on app design and adoption status (Table 7) showed that there is a main effect of adoption status ( $F_{161,1}=9.51$ ;  $P=.002$ ) and an interaction between app design and adoption status ( $F_{161,1}=4.03$ ;  $P=.046$ ).

**Figure 8.** Mean ratings of perceived persuasiveness of the diagnosis report interface for COVID Alert adopters and nonadopters. Horizontal bar represents the overall mean value of the construct for each user group. Vertical bars represent 95% CIs. C: control design; P: persuasive design.



Owing to the interaction between adoption status and app design, we carried out a further 1-way ANOVA at each level of each factor as shown in Table 8. The results showed that there was a main effect of app design within the adopter group ( $F_{64,1}=8.00$ ;  $P=.006$ ), with a medium effect size ( $\eta_p^2=0.11$ ). In other words, the persuasive design (mean 6.00, SD 0.97) had a significantly

higher mean value for perceived persuasiveness than the control design (mean 5.03, SD 1.22). Moreover, adoption status had a main effect regarding the persuasive design, with a near large effect size ( $F_{76,1}=11.10$ ;  $P=.001$ ;  $\eta_p^2=0.13$ ). In other words, adopters (mean 6.00, SD 0.97) significantly rated the perceived persuasiveness of the persuasive design higher than that of nonadopters (mean 4.61, SD 1.84).

**Table 7.** Repeated-measure ANOVA based on app design, adoption status, and perceived persuasiveness indicator for the diagnosis report interface.

	Adoption status	App design×adoption status
<i>Df Res</i> <sup>a</sup>	161	161
<i>F (df)</i>	9.51 (1)	4.03 (1)
<i>P value</i>	.002	.046

<sup>a</sup>*Df Res*: degree of freedom residual.

**Table 8.** Further 1-way ANOVA for the perceived persuasiveness of the diagnosis report interface at each level of adoption status and app design (small effect size:  $\eta_p^2=0.01$ ; medium effect size:  $\eta_p^2=0.06$ ; larger effect size:  $\eta_p^2=0.14$ ) [76].

	One-way ANOVA for each app design		App design effect
	C3 <sup>a</sup>	P3 <sup>b</sup>	
<b>One-way ANOVA within each adoption status</b>			
Adopter	5.03	6.00	$F_{64,1}=8.00$ ; $P=.006$ ; $\eta_p^2=0.11$
Nonadopter	4.77	4.61	$F_{97,1}=0.00$ ; $P=.99$
Adoption effect	$F_{85,1}=0.56$ ; $P=.46$	$F_{76,1}=11.10$ ; $P<.001$ ; $\eta_p^2=0.13$	N/A <sup>c</sup>

<sup>a</sup>C: control design.

<sup>b</sup>P: persuasive design.

### Two-Way ANOVA for Each App Design

In this section, due to the 3-way interaction in Table 8, we conducted a 2-way ANOVA based on adoption status and interface for each of the 3 interfaces.

#### Control Design

Table 9 presents the 2-way ANOVA based on the adoption status and interface for the control design. The results show a main effect of adoption status ( $F_{252,1}=20.00$ ;  $P<.001$ ) and an interaction between adoption status and interface ( $F_{252,2}=3.45$ ;  $P=.03$ ).

Owing to the interaction between interface and adoption status, we carried out a further 1-way ANOVA at each level of each factor as shown in Table 10. The results show that there is a

main effect of adoption status with regard to the no-exposure status interface ( $F_{79,1}=20.41$ ;  $P<.001$ ;  $\eta_p^2=0.21$ ) and the exposure status interface ( $F_{88,1}=21.85$ ;  $P<.001$ ;  $\eta_p^2=0.20$ ), with a large effect size. In both interfaces, the mean value of perceived persuasiveness was significantly higher for the adopter group than for the nonadopter group. Moreover, there was a main effect of interface within adopters, ( $F_{99,2}=6.33$ ;  $P=.003$ ), with a near large effect size ( $\eta_p^2=0.13$ ), which made us carry out a further pairwise comparison. The results showed that the mean values of perceived persuasiveness for the no-exposure status interface (mean 5.87, SD 1.20) and exposure status interface (mean 6.12, SD 1.01) were significantly higher than those of the diagnosis report interface (mean 5.03, SD 1.22) at  $P=.04$  and  $P=.003$ , respectively.

**Table 9.** Two-way ANOVA based on adoption status and interface for the control design.

	Adoption status	Interface×adoption status
<i>Df Res</i> <sup>a</sup>	252	252
<i>F (df)</i>	20.00 (1)	3.45 (2)
<i>P value</i>	<.001	.03

<sup>a</sup>*Df Res*: degree of freedom residual.

**Table 10.** Further 1-way ANOVA for the perceived persuasiveness of the control design at each level of interface and adoption status (small effect size:  $\eta^2=0.01$ ; medium effect size:  $\eta^2=0.06$ ; larger effect size:  $\eta^2=0.14$ ) [76].

	One-way ANOVA for each interface			Interface effect
	No-exposure status	Exposure status	Diagnosis report	
<b>One-way ANOVA within each adoption status</b>				
Adopter	5.87	6.12	5.03	$F_{99,2}=6.33$ ; $P=.003$ ; $\eta_p^2=0.13$
Nonadopter	4.57	4.82	4.78	$F_{153,2}=0.92$ ; $P=.40$
Adoption effect	$F_{79,1}=20.41$ ; $P<.001$ ; $\eta_p^2=0.21$	$F_{88,1}=21.85$ ; $P<.001$ ; $\eta_p^2=0.20$	$F_{85,1}=0.56$ ; $P=.46$	N/A <sup>a</sup>

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

**Persuasive Design**

Table 11 shows the 2-way ANOVA based on adoption status and interface for persuasive design. The results show that there is a main effect of adoption status ( $F_{279,1}=4.96$ ;  $P=.03$ ;

$\eta_p^2=0.03$ ), with the mean value of perceived persuasiveness of the persuasive design being significantly higher for the adopter group (mean 5.69, SD 1.24) than for the nonadopter group (mean 5.01, SD 1.54). There is no interaction between adoption status and interface.

**Table 11.** Two-way ANOVA based on adoption status and interface for the persuasive design (small effect size:  $\eta^2=0.01$ ; medium effect size:  $\eta^2=0.06$ ; larger effect size:  $\eta^2=0.14$ ) [76].

	No-exposure status	Exposure status	Diagnosis report	Overall
Adopter	5.36	5.70	6.00	5.69
Nonadopter	5.37	5.05	4.61	5.01
Adoption effect	N/A <sup>a</sup>	N/A	N/A	$F_{279,1}=4.96$ ; $P=.03$ ; $\eta_p^2=0.03$

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

**Two-Way ANOVA for Each Adoption Status**

In this section, owing to the 3-way interaction in Table 3, we conducted a 2-way ANOVA based on app design and interface for each adoption status.

**Adopter Group**

We performed a 2-way ANOVA based on the app design and interface for the adopter group. The results showed that there was an interaction between app design and interface ( $F_{189,2}=6.73$ ;  $P=.001$ ). Owing to the interaction, we carried out a further 1-way ANOVA at each level of app design and

interface as shown in Table 12. The results show that there is a main effect of the interface with regard to the control design ( $F_{99,2}=6.33$ ;  $P=.003$ ;  $\eta_p^2=0.13$ ). There was also a main effect of app design with regard to the diagnosis report interface ( $F_{64,1}=8.00$ ;  $P=.006$ ;  $\eta_p^2=0.11$ ). Finally, there is a main effect of app design in the exposure status interface ( $F_{64,1}=4.31$ ;  $P=.04$ ;  $\eta_p^2=0.06$ ). Regarding the diagnosis report interface, the mean of perceived persuasiveness is significantly higher for the persuasive design than the control design. However, the reverse is true for the exposure status interface.

**Table 12.** Further 1-way ANOVA for adopters' perceived persuasiveness at each level of app design and interface (small effect size:  $\eta^2=0.01$ ; medium effect size:  $\eta^2=0.06$ ; larger effect size:  $\eta^2=0.14$ ) [76].

	One-way ANOVA for each interface			Interface effect
	No-exposure status	Exposure status	Diagnosis report	
<b>One-way ANOVA within each app design</b>				
Control design	5.87	6.12	5.03	$F_{99,2}=6.33$ ; $P=.002$ ; $\eta_p^2=0.13$
Persuasive design	5.36	5.70	6.00	$F_{90,2}=0.98$ ; $P=.38$
Adoption effect	$F_{61,1}=1.05$ ; $P=.31$	$F_{64,1}=4.3$ ; $P=.04$ ; $\eta_p^2=0.06$	$F_{64,1}=8.00$ ; $P=.006$ ; $\eta_p^2=0.11$	N/A <sup>a</sup>

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

### Nonadopter Group

Table 13 shows a 2-way ANOVA based on app design and interface for the nonadopter group. The results show that there

is a main effect of app design ( $F_{342,1}=5.62$ ;  $P=.02$ ;  $\eta_p^2=0.02$ ), with a small effect size and persuasive design (mean 5.01, SD 1.54) having a significantly higher mean value of perceived persuasiveness than the control design (mean 4.72, SD 1.25).

**Table 13.** Two-way ANOVA based on app design and interface for the nonadopter group (small effect size:  $\eta_p^2=0.01$ ; medium effect size:  $\eta_p^2=0.06$ ; larger effect size:  $\eta_p^2=0.14$ ) [76].

	No-exposure status	Exposure status	Diagnosis report	Overall
Control design	4.56	4.82	4.77	4.72
Persuasive design	5.37	5.05	4.61	5.01
App design effect	N/A <sup>a</sup>	N/A	N/A	$F_{342,1}=5.62$ ; $P=.02$ ; $\eta_p^2=0.02$

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

### Nonadopters' Willingness to Download the COVID Alert App

This section addresses the fourth set of hypotheses (H4). Figure 9 shows the percentages of nonadopters in each of the 6 groups who were willing to download the COVID Alert app from the Apple or Google store after completing the survey. The question they responded to was *Now that I know about the COVID Alert app as the Government of Canada's official exposure notification app, I will download it from the Apple/Google store to slow down the spread of the coronavirus*. This question was targeted only at nonadopters in the survey. Overall, the percentage of nonadopters willing to download the app from the app store was higher for the persuasive design (37/64, 58%) than for the control design (24/52, 46%).

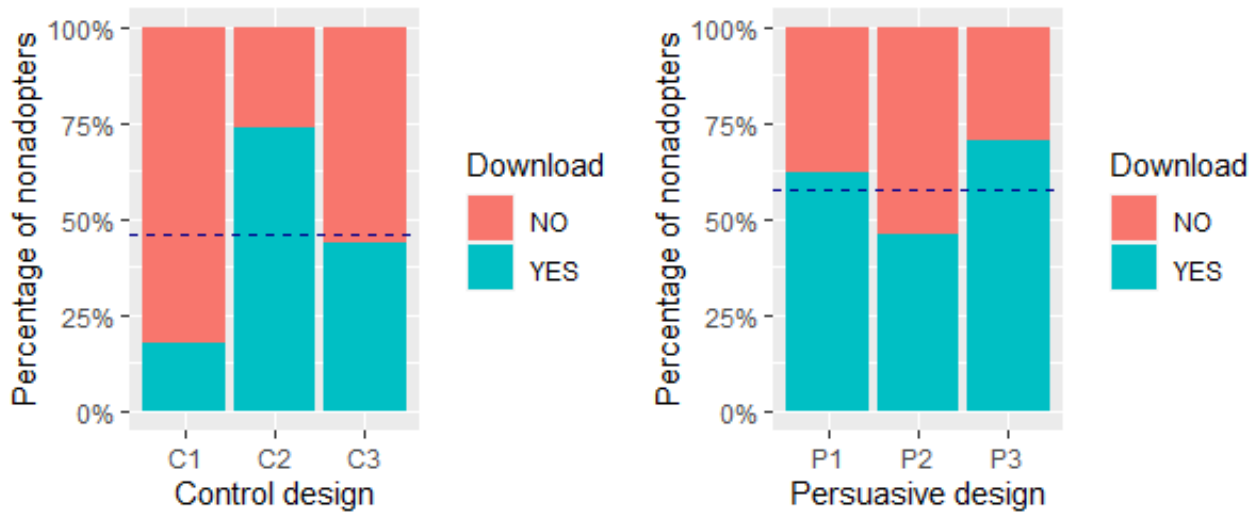
For the no-exposure status interface, the percentage of *yes* responses was higher for P1 (13/21, 62%) than for C1 (3/17, 18%). Similarly, for the diagnosis report interface, the percentage of *yes* responses was higher for P3 (12/17, 71%) than for C3 (7/16, 44%). However, for the exposure status interface, the percentage of *yes* responses was higher for C2 (14/19, 74%) than for P2 (12/26, 46%). To investigate the statistically significant difference between each pair of interface designs (C1 vs P1, C2 vs P2, and C3 vs P3), we carried out a chi-square test as shown in Table 14. Overall, the test showed

a significant difference between at least one of the pairs ( $\chi^2_5=88.01$ ;  $P<.001$ ). Next, for the 6 user interfaces, we carried out a post hoc pairwise chi-square test using the *pairwiseNominalIndependence* function from the *rcompanion* package in R, and the Benjamini–Hochberg false discovery rate method of correction for multiple comparison errors [77]. The test showed that the persuasive and control designs for each of the 3 pairs of interfaces were significantly different ( $P<.001$ ). We also computed the effect size ( $\phi$ ) based on a 2x2 contingency table for each type of interface as shown in Table 14. We used the *chisq\_to\_phi* function from the *effectsize* package [78] to compute the size of the effect of persuasive design on each interface. The result of the computation showed that the effect size of persuasive design for the 3 interfaces is large ( $\phi\geq 0.50$ ), with that regarding the no-exposure status interface being the highest ( $\phi=1.01$ ).

It is noteworthy that C2 accruing more *yes* responses (14/19, 74%) than P2 (12/26, 46%), coupled with the nonsignificant difference between the perceived persuasiveness of both interfaces ( $P=0.53$ , Table 6) indicates that the nonadopters prefer the control design of the exposure status interface over the persuasive design. Altogether, P1, C2, and P3 are preferred over C1, P2, and C3. Figure 10 shows the overall percentage of *yes* responses for each set of interfaces, with the former (39/57, 68%) exceeding the latter (22/59, 37%) by >30%.



**Figure 9.** Percentages of nonadopters willing to download the COVID Alert app. Horizontal bar represents the overall percentage of nonadopters in each app design who were willing to download the app. C: control design; P: persuasive design.



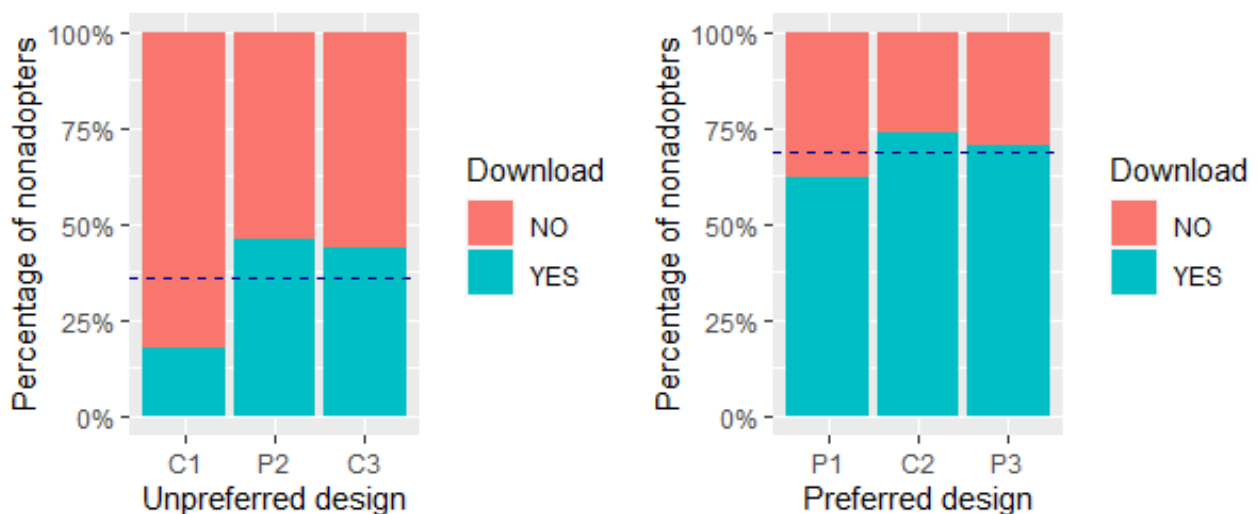
**Table 14.** Chi-square and pairwise comparison tests for nonadopters willing to download the COVID Alert app based on Benjamini–Hochberg false discovery rate method of correction for multiple comparison errors (small effect size:  $\phi=0.1$ ; medium effect size:  $\phi=0.3$ ; larger effect size:  $\phi=0.5$ ) [68,78,79].

	No-exposure status interface		Exposure status interface		Diagnosis report interface		P value	Chi-square (df)										
	C1 <sup>a</sup>	P1 <sup>b</sup>	Comparison	C2	P2	Comparison			C3	P3	Comparison							
<b>Willing to download the COVID Alert app</b>			$P<.001$ ; $\phi=1.01$			$P<.001$ ; $\phi=0.57$			$P<.001$ ; $\phi=0.64$									
Yes (%)	17.65	61.90		73.68	46.15				43.75	70.59								
No (%)	82.35	38.10		26.32	53.85				56.25	29.41								
Difference (%)	-64.70	+23.80		+47.36	-7.70				-12.50	+41.48								

<sup>a</sup>C: control design.

<sup>b</sup>P: persuasive design.

**Figure 10.** Percentages of nonadopters willing to download the COVID Alert app, with C2 and P2 switched to realize the preferred set of interfaces on the right. Horizontal bar represents the overall percentage of nonadopters in each app design who were willing to download the app. C: control design; P: persuasive design.



## Discussion

### Principal Findings

In this section, we discuss our findings in the context of our hypotheses. For ease of reference, we summarize the key findings in Table 15. Overall, 83% (10/12) of hypotheses were fully or partially supported by the empirical data and analysis. By partial support, we mean that the hypothesis in question is only supported with regard to one of the adoption groups (adopters or nonadopters) or app designs (persuasive or control). Overall, the study reveals that adopters found the COVID Alert

app, regardless of app design and use case, more persuasive than nonadopters (H3a, H3b, and H3c). Second, the study reveals that the persuasive design is more likely to be effective than the control design in motivating nonadopters to adopt exposure notification apps (H2a, H4a, and H4c) and adopters to report their COVID-19 diagnoses (H2c). In other words, our findings suggest that contact tracing apps are more likely to be effective if they are designed as persuasive technologies, particularly by incorporating self-monitoring that helps users track number of daily contacts and duration of exposure, and social learning that motivates users to report their COVID-19 diagnosis through social pressure.

**Table 15.** Summary of the validation of hypotheses.

Hypothesis (H) number	Hypothesis	Remark
H1a	The higher the perceived persuasiveness of the no-exposure status interface in the app store, the more likely users will download the COVID Alert app.	Supported
H1b	The higher the perceived persuasiveness of the exposure status interface in the app store, the more likely users will download the COVID Alert app.	Supported
H1c	The higher the perceived persuasiveness of the diagnosis report interface in the app store, the more likely users will download the COVID Alert app.	Supported
H2a	The perceived persuasiveness of the persuasive design of the no-exposure status interface will be higher than that of the control design.	Supported among nonadopters only
H2b	The perceived persuasiveness of the persuasive design of the exposure status interface will be higher than that of the control design.	Not supported
H2c	The perceived persuasiveness of the persuasive design of the diagnosis report interface will be higher than that of the control design.	Supported among adopters only
H3a	Adopters are more likely to perceive the no-exposure status interface to be persuasive than nonadopters.	Supported overall and particularly regarding the control design
H3b	Adopters are more likely to perceive the exposure status interface to be persuasive than nonadopters.	Supported overall
H3c	Adopters are more likely to perceive the diagnosis report interface to be persuasive than nonadopters.	Supported overall and particularly, regarding the persuasive design
H4a	Nonadopters who viewed the persuasive design of the no-exposure status interface are more likely to adopt the COVID Alert app than those who viewed the control design.	Supported
H4b	Nonadopters who viewed the persuasive design of the exposure status interface are more likely to adopt the COVID Alert app than those who viewed the control design.	Not supported: the reverse was the case
H4c	Nonadopters who viewed the persuasive design of the diagnosis report interface are more likely to adopt the COVID Alert app than those who viewed the control design.	Supported

### Relationship Between Perceived Persuasiveness and Willingness to Download the COVID Alert App

Our path models supported the first 3 hypotheses. Regarding each user interface, we found that the relationship between perceived persuasiveness and willingness to download the app is significant. The relationship was strongest for the no-exposure status interface ( $\beta=.68$ ;  $P<.001$ ), followed by the exposure status interface ( $\beta=.67$ ;  $P<.001$ ) and the diagnosis report interface ( $\beta=.47$ ;  $P=.04$ ). On the basis of the multigroup analysis, there was no statistically significant difference between each pair of path coefficients. Hence, the first set of hypotheses, *the higher the perceived persuasiveness of each interface, the more likely users will download the COVID Alert app (H1a, H1b, and H1c)*,

is supported. This finding is consistent with the finding by Oyibo and Vassileva [57] in the physical activity domain. The authors found that the higher users perceive a fitness app to be persuasive, the higher their intention to use the app to motivate behavior change.

Moreover, the 3 models have an acceptably large goodness of fit (GOF), which shows how well the model fits the data. The GOF for the no-exposure and exposure status interfaces was  $>60\%$ , and that of the diagnosis report interface was  $38\%$ . As stated by Hussain et al [80], a GOF for  $36\%$  is regarded as large. Moreover, perceived persuasiveness in the models regarding the no-exposure and exposure status interfaces explains at least  $40\%$  of the variance in respondents' willingness to download the app. However, in the model for the diagnosis report interface,

only 20% of the target construct was explained by perceived persuasiveness. More than 60% is regarded as a high explanation of the variance of the target construct and <30% is regarded as a low explanation [74]. Therefore, the variance in willingness to download the app explained for the no-exposure and exposure status interfaces is medium and that for the diagnosis report interface is small. These findings, which correlate with the magnitude and significance of the relationships between perceived persuasiveness and willingness to download the app (Figure 5), indicate that self-monitoring, which the no-exposure and exposure status interfaces support, is more likely to motivate nonadopters to download the app than the diagnosis reporting feature of the app. This finding may not be surprising given that notification of COVID-19 exposure and monitoring of exposure levels tend to benefit the user personally, whereas diagnosis reporting tends to benefit the community. This plausible explanation is reflected in the mean ratings of the perceived persuasiveness of the 2 interfaces by the 2 groups. For the nonadopters, the overall perceived persuasiveness of the user interfaces (Figures 5-7) is numerically higher for the no-exposure status interface (mean 5.01, SD 1.54) and the exposure status interface (mean 4.96, SD 1.43) than for the diagnosis report interface (mean 4.69, SD 1.54). Similarly, for the adopters, the perceived persuasiveness of the control interfaces (Table 10) was significantly higher for the no-exposure status interface (mean 5.87, SD 1.20) and the exposure status interface (mean 6.12, SD 1.01) than for the diagnosis report interface (mean 5.03, SD 1.22).

### App Design Effect on Perceived Persuasiveness

In this section, we discuss the effect of app design (persuasive vs control) on the perceived persuasiveness of each of the 3 user interfaces.

#### No-Exposure Status Interface

Regarding the perceived persuasiveness of the no-exposure status interface, we found an interaction between app design and adoption status (Table 4). Among the adopters, the perceived persuasiveness of the control design and that of the persuasive design did not differ significantly ( $P=.31$ , Table 5). However, among nonadopters, the perceived persuasiveness of the persuasive design (mean 5.37, SD 1.30) was significantly higher than that of the control design (mean 4.57, SD 1.19). The effect size of the mean difference between the 2 app designs was medium ( $\eta_p^2=0.10$ ). Therefore, the fourth hypothesis (H2a), *the perceived persuasiveness of the persuasive design of the no-exposure status interface will be higher than that of the control design*, is validated for nonadopters. This finding is an indication that although the app design does not matter among adopters, it does matter among nonadopters. This implies that nonadopters are more likely to adopt the persuasive version of the no-exposure status interface (with self-monitoring features) than the control version (without self-monitoring features).

It is noteworthy that, among nonadopters, although demographic variables may confound the validation of H2a, gender is less likely to do so. This is because the gender-based distributions of the nonadopter group that evaluated the control design (C1) and that of the nonadopter group that evaluated the persuasive

design (P1) were very similar. As shown in Multimedia Appendix 2, a total of 75% (12/16) of the C1 adopter group were men, and 25% (4/16) were women. Similarly, 71% (15/21) of the P1 adopter group were men, and 29% (6/21) were women. However, the percentage distributions based on age and education for the C1 and P1 nonadopters were different. For example, 24% (5/21) of the P1 nonadopter group were aged <25 years, whereas 0% (0/15) of the C1 nonadopter group were aged <25 years. Moreover, in the P1 nonadopter group, 25% (5/20) had high school qualification, compared with only 6% (1/17) in the C1 nonadopter group. One plausible explanation for the higher percentage of participants with lower education in the P1 nonadopter group than in the C1 nonadopter group is that the former group had a higher percentage of younger participants aged <25 years. Hence, in future analyses, we hope to investigate the effect of age and education on the significant difference between the P1 and C1 nonadopter groups, which may partly account for the perception of P1 as more persuasive than C1.

#### Exposure Status Interface

Regarding the perceived persuasiveness of the exposure status interface, we did not find an effect of app design on perceived persuasiveness (Table 6). Hence, the fifth hypothesis (H2b), *the perceived persuasiveness of the persuasive design of the exposure status interface will be higher than that of the control design*, was not validated. One plausible reason why the persuasive design is not perceived as more persuasive than the control design by either adopters or nonadopters is that the information displayed on the exposure status interface is historical. In other words, the displayed information on the exposure status interface is the total sum of exposure levels over a 14-day period. This cumulative information is less transparent and unlike that of the no-exposure status interface where the displayed exposure level is for each day. Hence, the persuasive version of the no-exposure status interface, which displays daily exposure levels, was perceived as more persuasive than the control version by the nonadopter group as shown in (Table 5).

#### Diagnosis Report Interface

Regarding the perceived persuasiveness of the diagnosis report interface, we found an interaction between app design and adoption status (Table 7). Among nonadopters, the perceived persuasiveness of the persuasive design and that of the control design did not differ significantly ( $P=.99$ , Table 8). However, among adopters, they differed significantly ( $P=.006$ ). Specifically, adopters perceived the persuasive design (mean 6.00, SD 0.97) to be more persuasive than the control design (mean 5.03, SD 1.22). The effect size of the mean difference between the 2 app designs was medium ( $\eta_p^2=0.11$ ). Therefore, the sixth hypothesis (H2c), *the perceived persuasiveness of the persuasive design of the diagnosis report interface will be higher than that of the control design*, is validated for adopters. A plausible explanation for this finding is that having used the control design of the COVID Alert app, the adopters are likely to find the persuasive design, which incorporates social learning, more persuasive. The additional message puts the user under social pressure to follow suit, ie, join other concerned individuals who have reported their diagnosis so that exposed contacts can

be notified and take the necessary safety measures to reduce the spread of the virus. The feeling of social pressure to report their COVID-19 diagnosis, fostered by the persuasive design, can be likened to the obligation and social pressure that the adopters must have felt upon the clarion call from the government and public health authorities for mass adoption to flatten the curve. However, for the nonadopters, the socially pressuring message in the persuasive design makes no significant difference compared with the control design ( $P=.99$ ). One plausible explanation for the nonsignificant difference between both app designs among the nonadopter group is that, compared with adopters, they are less responsive to socially oriented messages, be it from the government, public health authorities, or the app. Hence, we see that the adopters in real life adopted COVID Alert owing to the clarion call from the government and public health authorities, whereas the nonadopters did not.

It is noteworthy that, among adopters, although demographic variables may confound the validation of H2c, gender and education were less likely. This is because the percentage distribution of the adopter group that evaluated the persuasive design (P3) based on gender and education and that of the adopter group that evaluated the control design (C3) look similar (Multimedia Appendix 2). For example, regarding gender, 67% (6/9) of the adopter participants who evaluated C3 were men, and 33% (3/9) were women. The same percentage distribution applies to the adopter participants who evaluated P3: 67% (8/12) were men, and 33% (4/12) were women. Similarly, regarding education, 23% (3/13) of the C3 adopters vs 22% (2/9) of the P3 adopters participants had a high school qualification, 62% (8/13) vs 56% (5/9) had a bachelor's degree, and 15% (2/13) vs 22% (2/9) had a master's degree. However, the percentage distributions based on age and smartphone use experience for the C3 and P3 adopter groups were different. For example, 100% (13/13) of the participants in the C3 adopter group were aged <45 years compared with 78% (7/9) in the P3 adopter group. Moreover, 85% (11/13) of the C3 adopter group had >5 years of experience, compared with 100% (8/8) of the P3 adopter group. One plausible explanation for the higher percentage of participants with more years of smartphone use experience in the P3 adopter group than in the C3 adopter group is that the former group had a higher percentage of older participants. Hence, in future analyses, we hope to uncover the effect of age and smartphone use experience on the significant difference between the P3 and C3 adopter groups, which may partly account for the perception of P3 as more persuasive than C3.

### Adoption Effect on Perceived Persuasiveness

In this section, we discuss the effect of adoption status (adopter vs nonadopter) on the perceived persuasiveness of each of the 3 user interfaces.

#### No-Exposure Status Interface

Regarding the perceived persuasiveness of the no-exposure status interface, we found an interaction between the adoption status and app design (Table 4). Regarding persuasive design (Table 5), there was no significant difference between adopters and nonadopters ( $P=.99$ ). However, regarding the control design, there was an adoption status effect, with adopters (mean 5.87,

SD 1.20) perceiving the user interface to be more persuasive than nonadopters (mean 4.57, SD 1.19). The effect size of the mean difference between the adoption statuses was large ( $\eta_p^2=0.21$ ). Therefore, the seventh hypothesis (H3a), *adopters are more likely to perceive the no-exposure status interface to be persuasive than nonadopters*, is validated for the control design. A plausible explanation for this finding is that, overall, the COVID Alert adopters are more concerned with the social benefit of using contact tracing apps to curb the spread of the coronavirus than nonadopters. This explains why they are among the early adopters of the app compared with the nonadopters. Hence, it stands to reason that the adopters are more likely to perceive the COVID Alert app that they are currently using to be persuasive than the nonadopters, who are yet to adopt the app.

It is noteworthy that demographic variables such as gender and smartphone use experience may confound the validation of H3a. The reason is that the distribution of the adopter and nonadopter groups that evaluated the control design (C1) based on 3 demographic factors differs one way or the other. As shown in Multimedia Appendix 2, a total of 40% (4/10) of the C1 adopter participants were men, compared with 75% (12/16) of the C1 nonadopter group. Moreover, based on smartphone use experience, we had a higher percentage of participants with lower and higher experience in the C1 nonadopter group than in the C1 adopter group. As shown in Multimedia Appendix 2, a total of 18% (3/17) of the C1 nonadopter group had <6 years of experience and 12% (2/17) had >20 years of experience, compared with 0% (0/10) of both experience levels in the C1 adopter group. Hence, in future analyses, we hope to investigate the effect of gender and smartphone use experience on the significant difference between the C1 adopter and nonadopter groups, which may partly account for the perception of C1 by the former group as more persuasive than the latter group.

#### Exposure Status Interface

Regarding the exposure status interface, our ANOVA showed that adoption had a main effect (Table 6), with adopters perceiving the interface to be more persuasive (mean 5.91, SD 1.01) than nonadopters (mean 4.96, SD 1.43). The effect size of the mean difference between adoption status was medium ( $\eta_p^2=0.09$ ). Hence, the eighth hypothesis (H3b), *adopters are more likely to perceive the exposure status interface to be persuasive than nonadopters*, is validated regardless of the app design. A plausible explanation for this finding is that, compared with the nonadopters, the adopters are more likely to be committed to the social cause of curbing the spread of the coronavirus and thus are more likely to be persuaded to use the COVID Alert app. This explains why they installed the COVID Alert app in the first place and are using it to track their exposure status (at the time of the study).

#### Diagnosis Report Interface

Regarding the diagnosis report interface (Table 7), we found an interaction between app design and adoption status regarding the perceived persuasiveness of the interface. Regarding the control design (Table 8), there was no significant difference between adopters and nonadopters ( $P=.46$ ). However, regarding

the persuasive design, there is an adoption effect, with adopters (mean 6.00, SD 0.97) perceiving the user interface to be more persuasive than nonadopters (mean 4.61, SD 1.84). The effect size of the mean difference between the 2 groups was near large ( $\eta_p^2=0.13$ ). Therefore, the ninth hypothesis (H3c), *adopters are more likely to perceive the diagnosis report interface to be persuasive than nonadopters*, is validated with regard to the persuasive design. A plausible explanation for this finding is that adopters, overall, are more motivated and concerned about the social obligation to curb the spread of the coronavirus using contact tracing apps than the nonadopters, as discussed earlier in Section 5.2 Diagnosis Report Interface. In fact, not only did adopters find the persuasive design significantly more persuasive (mean 6.00, SD 0.97) than nonadopters (mean 4.61, SD 1.84) they also found it more persuasive than the control design (mean 5.03, SD 1.22). However, this is not the case for nonadopters, who did not perceive the persuasiveness of the persuasive design (mean 4.61, SD 1.84) significantly different from that of the control design (mean 4.77, SD 1.21).

It is noteworthy that apart from adoption status, demographic variables such as gender, age, education, and smartphone use experience may partly account for the significant difference between the adopter group and the nonadopter group that evaluated P3 (H3c). For example, as shown in [Multimedia Appendix 2](#), two-thirds of the P3 adopter group were men (6/9, 67%), while one-third were men in the P3 nonadopter group (6/17, 35%). Moreover, 41% (7/17) of the P3 nonadopter group had 1 to 5 years of smartphone use experience, whereas 100% (8/8) of the participants in the P3 adopter group had >5 years of experience. Hence, in future analyses, we hope to investigate the effect of gender, smartphone use experience, and other demographic factors on the significant difference between the P3 adopter and nonadopter groups. The demographic factors may partly account for the perception of P3 by the adopter group as more persuasive than the nonadopter group. Research questions such as (1) *Are people more likely to perceive the persuasive interfaces (eg, P3) as persuasive with increase in smartphone use experience (as the percentage distribution in Multimedia Appendix 2 seems to suggest)* will be addressed and (2) *Are males more likely to perceive the persuasive interfaces (eg, P3) as persuasive than females (as the percentage distribution in Multimedia Appendix 2 seems to suggest)* will be addressed.

### Adoption Effect on Willingness to Download the COVID Alert App

Among the nonadopters, the chi-square tests regarding willingness to download the COVID Alert app show that there is an effect of user interface. This led us to carry out post hoc pairwise comparisons to uncover the effect of app design. Regarding the no-exposure status interface, the pairwise comparison shows that the size of the effect of the persuasive design is large ([Table 14](#)). This indicates that the group that viewed the persuasive design (13/21, 62%) was more willing to download the app than the group that viewed the control design (3/17, 18%). Hence, the tenth hypothesis (H4a), *nonadopters who viewed the persuasive design of the no-exposure status interface are more likely to adopt the COVID*

*Alert app than those who viewed the control design*, is validated. This finding was replicated with regard to the diagnosis report interface. Those who viewed the persuasive design (12/17, 71%) were more willing to download the app than those who viewed the control design (7/16, 44%). Thus, the twelfth hypothesis (H4c), *nonadopters who viewed the persuasive design of the diagnosis report interface are more likely to adopt the COVID Alert app than those who viewed the control design*, is validated. The validation of H4a and H4c corroborates the findings in [Table 13](#): among the nonadopter group, the overall perceived persuasiveness of the persuasive designs (mean 5.01, SD 1.54) is significantly higher than that of the control designs (mean 4.72, SD 1.25).

However, although the effect size tests for P1 and P3 showed that the persuasive designs were more likely to be downloaded by the participants than the control designs (C1 and C3), the reverse was true for C2 and P2. The effect size test for the exposure status interface indicated that the 11th hypothesis (H4a), *nonadopters who viewed the persuasive design of the exposure status interface (P2) are more likely to adopt the COVID Alert app than those who viewed the control design (C2)*, was not validated. Specifically, only 46% (12/26) of those who viewed the persuasive design were willing to download the app, compared with 74% (14/19) of those who viewed the control design. This finding is counterintuitive, given that the nonadopters who viewed the other 2 persuasive designs (P1 and P3) were more willing to download the app than those who viewed the control designs (C1 and C3). Although the finding is counterintuitive, it may not be far-fetched given that it aligns with the finding that among adopters ([Table 12](#)), the perceived persuasiveness of the control exposure status interface (mean 6.12, SD 1.01) is significantly higher than that of its persuasive version (mean 5.70, SD 1.02). One plausible explanation for this counterintuitive finding is the idea that the app keeps a record of the user's total number of contacts and exposure minutes within the last 14 days ([Figure 2](#)), which, in the context of privacy, users may not like. The historical record displayed by the app may be perceived as individual surveillance [81]. Second, it has the potential to reveal the individual from whom the user contracted the virus if the total number of contacts over the 14-day rolling period was small. This may partly explain the poor performance of the persuasive version of the exposure status interface among adopters and nonadopters. Another plausible explanation for the counterintuitive finding is the relatively high hypothetical statistics presented in the P2 interface, which may be far from reality. In other words, viewing relatively high number of contacts and exposure time within the last 14 days (75 persons and 212 minutes) might have made some of them feel very uncomfortable and even doubtful. The reason for this assertion is that one would have expected the percentage of the P2 group of participants willing to download the app to be much higher given that (1) they could view the cumulative sum of their contacts and exposure minutes, which is an added value and (2) the P1 and P3 groups, who viewed the persuasive designs, were more willing to download the app than the C1 and C3 groups, respectively, who viewed the control designs. In other words, the hypothetical numbers might have been significantly higher than what the P2 group expected in a real-life setting; for example, based on their actual social

distancing behavior, such as staying at and working from home. This might have caused cognitive dissonance, thereby making the P2 group doubt the accuracy of the app, which might have negatively affected their willingness to download it. In future work, we will investigate how the number of contacts and exposure time displayed in the exposure status interface influence its perceived persuasiveness and participants' willingness to download the app.

Moreover, in future work, we will investigate the possible effects of demographic factors such as gender, age, education, and smartphone use experience on the willingness to download the app. This might help explain why the group that viewed the control design of the exposure status interface was more willing to download the app than the group that viewed the persuasive design. However, by merely inspecting the percentage demographic distribution for the C2 and P2 nonadopter groups of participants based on all 4 demographic factors, there seems to be little to no difference between the 2 groups ([Multimedia Appendix 2](#)). For example, regarding gender, 53% (10/19) of the C2 nonadopter group compared with 62% (16/26) of the P2 nonadopter group were men. Second, regarding education, 16% (3/19) of the C2 nonadopters vs 23% (6/26) of the P2 nonadopters had a high school qualification, 68% (13/19) vs 54% (14/26) had a bachelor's degree, and 11% (2/19) vs 19% (5/26) had a master's degree. The demographic similarities between both groups led us to the question *Apart from demographic variables, what else could possibly account for the difference between the C2 and P2 nonadopter groups in terms of their willingness to download the COVID Alert app?* The analysis of the qualitative data collected in this study and investigation of the effect of the total exposure levels displayed on the exposure status interface, in future work, can help answer this research question and gain more insights.

### Summary of Main Findings

We have shown that exposure notification apps can be designed as persuasive technologies to make them more effective in motivating behavior change. Our results revealed that exposure notification apps are more likely to be adopted and effective if they incorporate persuasive features such as self-monitoring and social learning. Our key findings can be summarized as follows:

1. Nonadopters find the persuasive design of the no-exposure interface of an exposure notification app to be more persuasive than the control design.
2. Nonadopters are more willing to download an exposure notification app with a persuasive design of the no-exposure status and diagnosis report interfaces than one with a control design.
3. Nonadopters are more willing to download an exposure notification app with a control design for the exposure status interface than one with a persuasive design.
4. Adopters are more likely to be motivated to report their COVID-19 diagnosis by the persuasive design of the diagnosis report interface than by the control design.
5. Adopters perceive the control design of the no-exposure and exposure status interfaces as more persuasive than the control design of the diagnosis report interface.

6. Adopters find an exposure notification app more persuasive than nonadopters.
7. Equipping only the no-exposure status and diagnosis report interfaces with self-monitoring and social learning, respectively, can increase adoption among nonadopters by >30%.

### Recommendations and Future Work

On the basis of the overall findings from [Figure 9](#), a total of 58% (37/64 nonadopters) who viewed the persuasive designs were more willing to download the app from the app stores than 46% (24/52 nonadopters) who viewed the control designs. In other words, the percentage of nonadopters willing to download it from app stores increased by >10% owing to the incorporation of persuasive features into the COVID Alert app. More importantly, incorporating persuasive features into the no-exposure status interface and diagnosis report interface only has the potential to increase adoption by >30%. The exposure status interface aside, two-thirds (25/38 nonadopters) who viewed the persuasive designs were willing to download it compared with one-third (10/33 nonadopters) who viewed the control designs. This finding, together with the validation of most of the hypotheses, indicates that overall, the persuasive design of an exposure notification app is more likely to be adopted and effective than the control design. Hence, we recommend that exposure notification app sponsors work toward incorporating persuasive features such as self-monitoring and social learning into future iterations to increase adoption and user experience and make them more effective in curbing the spread of COVID-19. However, because of privacy concerns (the possibility of knowing the person from whom the user contracted the virus), displaying the total number of contacts within the last 14 days of exposure may not be advisable for the exposure status interface. In future studies, this recommendation should be investigated further. Moreover, the potential effectiveness of the other persuasive features identified in our conceptual paper (tailoring, personalization, expertise, trustworthiness, authority, praise, reward, etc) [18] should be investigated as well; for example, how would praising or rewarding the user one way or the other for uploading their one-time COVID-19 diagnosis key influence their continued use of the app or their intention to report their future diagnosis if they test positive again?

### Contributions

This study is the first to conduct research of this nature (designing contact tracing apps as persuasive technologies), using an actual exposure notification app currently being used by Canadian residents (COVID Alert app) as proof of concept. In this study, we made several contributions to knowledge regarding the persuasive design of exposure notification apps to make them more effective in curbing the spread of COVID-19. We identified and presented 3 key user interfaces (no-exposure status, exposure status, and diagnosis report). Researchers can adopt these interfaces as a basis for future research on exposure notification apps, not only for the current COVID-19 pandemic but also for other epidemics and pandemics in the future that may require exposure notification apps. Moreover, designers can work toward improving the

design of exposure notification apps by incorporating persuasive features, such as self-monitoring and social learning, which we showed to be effective in the no-exposure status interface and diagnosis report interface, respectively. Finally, we showed empirically that the persuasive design of these 2 interfaces has the potential to increase adoption among nonadopters by >30%.

### Limitations

This study has limitations. The first limitation is the sample size. We only had an average of 30 participants in each of the 6 groups after data cleaning. Moreover, the participants recruited on the web (ie, on the Amazon Mechanical Turk platform) may not be representative of the entire Canadian population. For example, digital literacy and willingness to download the COVID Alert app may be higher among study participants recruited on the web [82]. This limitation may affect the generalization of the current findings to the entire Canadian population. Hence, there is a need for further research with larger sample sizes that are more representative of the Canadian population. This will help investigate how the current findings can be generalized to a larger Canadian population. Moreover, there is a need for similar research among national populations outside Canada to examine the generalizability of the findings to other countries with similar and different cultures. For example, in future, we hope to conduct a similar study among participants residing in the United States (which has an individualist culture similar to Canada's) and Nigeria (which has a collectivist culture different from Canada's). The second limitation of the study is the remuneration of the participants, which may have influenced their responses in some ways. The third limitation is that our findings are based on the Government of Canada's COVID Alert app, which is only targeted at the Canadian population. Hence, there is a need for further research on country-specific apps among other national populations to investigate how the current findings generalize across different countries and cultures. The fourth limitation of this study is that we did not, in our ANOVA, investigate the main and interaction effects of important demographic variables such as gender, age, education, and smartphone use experience on the findings,

although we did discuss their possible effects. The fifth limitation is that we did not investigate the entire range of persuasive strategies available from the PSD model. In addition to self-monitoring and social learning, other persuasive strategies may be instrumental in improving the persuasive design of contact tracing and exposure notification apps, with some being more likely to be effective in motivating certain health behaviors than others. Future work should address these limitations.

### Conclusions

Contact tracing and exposure notification apps may continue to be useful for a long time given the endemic potential of COVID-19 [83]. In this paper, we demonstrated that the persuasive design of an exposure notification app is more likely to be effective, using Canada's COVID Alert as proof of concept. First, we showed that nonadopters, through self-monitoring, prefer to track their daily exposure levels (number of contacts and exposure time) in addition to knowing their exposure status. However, they are not favorable toward knowing the total number of contacts and exposure time after being notified of possible exposure to the virus. This may be due to privacy concerns, which include the possibility of knowing the individual from whom one contracted the virus, if the total number of contacts over the 14-day rolling period is small. Second, we showed that adopters are more likely to be motivated to report their COVID-19 diagnosis using a persuasive design that supports social learning (knowing how many others have reported their diagnosis) than a control design. In summary, this study indicates that equipping the no-exposure status and diagnosis report interfaces of an exposure notification app with self-monitoring and social learning, respectively, can increase the percentage of nonadopters willing to download the app by >30%. In future work, we aim to investigate how demographic variables such as age, gender, and education moderate the effectiveness of persuasive features in exposure notification app design. We also look forward to investigating the relationship between perceived persuasiveness, on one hand, and intentions to install exposure notification apps, self-isolate, and report COVID-19 diagnosis, on the other hand.

---

### Acknowledgments

This project was funded by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2017-05310) and Cybersecurity and Privacy Institute, University of Waterloo.

---

### Conflicts of Interest

None declared.

---

#### Multimedia Appendix 1

Administration of app interfaces to participants.

[[DOCX File, 29 KB - formative\\_v6i9e34212\\_app1.docx](#) ]

---

#### Multimedia Appendix 2

Demographics of adopters and nonadopters based on gender, age, education, and smartphone use experience.

[[DOCX File, 82 KB - formative\\_v6i9e34212\\_app2.docx](#) ]

---

### References

1. Tracking COVID-19: contact tracing in the digital age. World Health Organization. 2020 Sep 9. URL: <https://www.who.int/news-room/feature-stories/detail/tracking-covid-19-contact-tracing-in-the-digital-age> [accessed 2021-01-10]
2. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. *Lancet Digit Health* 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30184-9](https://doi.org/10.1016/S2589-7500(20)30184-9)] [Medline: [32839755](https://pubmed.ncbi.nlm.nih.gov/32839755/)]
3. COVID-19 Delta Variant: Risk Assessment and Implications for Practice. Ontario Agency for Health Protection and Promotion (Public Health Ontario). 2021 Jul 23. URL: [https://www.publichealthontario.ca/-/media/documents/ncov/voc/2021/08/covid-19-delta-variant-risk-assessment-implications.pdf?sc\\_lang=en](https://www.publichealthontario.ca/-/media/documents/ncov/voc/2021/08/covid-19-delta-variant-risk-assessment-implications.pdf?sc_lang=en) [accessed 2022-06-04]
4. The Possibility of COVID-19 after Vaccination: Breakthrough Infections. Centers of Disease Control and Protection. 2021. URL: <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness/why-measure-effectiveness/breakthrough-cases.html> [accessed 2021-09-04]
5. Silberner J. Now Isn't the Time to Abandon Contact Tracing. *Wired*. 2021 Aug 26. URL: <https://www.wired.com/story/contact-tracing-delta-variant/> [accessed 2021-09-21]
6. Akinbi A, Forshaw M, Blinkhorn V. Contact tracing apps for the COVID-19 pandemic: a systematic literature review of challenges and future directions for neo-liberal societies. *Health Inf Sci Syst* 2021 Apr 13;9(1):18 [FREE Full text] [doi: [10.1007/s13755-021-00147-7](https://doi.org/10.1007/s13755-021-00147-7)] [Medline: [33868671](https://pubmed.ncbi.nlm.nih.gov/33868671/)]
7. Osmanliu E, Rafie E, Bédard S, Paquette J, Gore G, Pomey MP. Considerations for the design and implementation of COVID-19 contact tracing apps: scoping review. *JMIR Mhealth Uhealth* 2021 Jun 09;9(6):e27102 [FREE Full text] [doi: [10.2196/27102](https://doi.org/10.2196/27102)] [Medline: [34038376](https://pubmed.ncbi.nlm.nih.gov/34038376/)]
8. Sadasivan S. Illustrating with diversity and inclusion for the COVID Alert app. Canada Digital Service. 2020 Nov 26. URL: <https://digital.canada.ca/2020/11/26/illustrating-with-diversity-and-inclusion-for-the-covid-alert-app/> [accessed 2020-12-20]
9. Kukuk L. Analyzing adoption of COVID-19 contact tracing apps using UTAUT. University of Twente. 2020. URL: [http://essay.utwente.nl/81983/1/Kukuk\\_BA\\_EEMCS.pdf](http://essay.utwente.nl/81983/1/Kukuk_BA_EEMCS.pdf) [accessed 2022-06-04]
10. Turnbull S. COVID Alert app nears 3 million users, but only 514 positive test reports. *CTV News*. 2020 Sep 29. URL: <https://www.ctvnews.ca/health/coronavirus/covid-alert-app-nears-3-million-users-but-only-514-positive-test-reports-1.5125256> [accessed 2022-06-04]
11. O'Neill PH. No, coronavirus apps don't need 60% adoption to be effective. *MIT Technology Review*. 2020 Jun 5. URL: <https://www.technologyreview.com/2020/06/05/1002775/covid-apps-effective-at-less-than-60-percent-download/> [accessed 2021-02-16]
12. Sharma B. Evaluating the UX of the world's contact tracing apps. *UX Planet*. 2020 Oct 3. URL: <https://uxplanet.org/evaluating-the-ux-of-the-worlds-contact-tracing-apps-77187d8c0535> [accessed 2021-11-05]
13. Privacy review of the COVID Alert exposure notification application. Office of the Privacy Commissioner of Canada. 2020 Jul 31. URL: [https://priv.gc.ca/en/privacy-topics/health-genetic-and-other-body-information/health-emergencies/rev\\_covid-app/](https://priv.gc.ca/en/privacy-topics/health-genetic-and-other-body-information/health-emergencies/rev_covid-app/) [accessed 2021-11-05]
14. Haggart B. Canada's COVID Alert app is a case of tech-driven bad policy design. *The Conversation*. 2020 Aug 13. URL: <https://theconversation.com/canadas-covid-alert-app-is-a-case-of-tech-driven-bad-policy-design-144448> [accessed 2021-11-05]
15. Bahrain, Kuwait and Norway contact tracing apps among most dangerous for privacy. *Amnesty International*. 2020 Jun 16. URL: <https://www.amnesty.org/en/latest/news/2020/06/bahrain-kuwait-norway-contact-tracing-apps-danger-for-privacy/> [accessed 2021-11-05]
16. Li T, Cobb C, Yang JJ, Baviskar S, Agarwal Y, Li B, et al. What makes people install a COVID-19 contact-tracing app? Understanding the influence of app design and individual difference on contact-tracing app adoption intention. *Pervasive Mob Comput* 2021 Aug;75(C):101439. [doi: [10.1016/j.pmcj.2021.101439](https://doi.org/10.1016/j.pmcj.2021.101439)]
17. Cruz MM, Oliveira RS, Beltrão AP, Lopes PH, Viterbo J, Trevisan DG, et al. Assessing the level of acceptance of a crowdsourcing solution to monitor infectious diseases propagation. In: *Proceedings of the 2020 IEEE International Smart Cities Conference*. 2020 Presented at: ISC2 '20; September 28-October 1, 2020; Piscataway, NJ, USA p. 1-8 URL: <https://doi.org/10.1109/ISC251055.2020.9239069> [doi: [10.1109/isc251055.2020.9239069](https://doi.org/10.1109/isc251055.2020.9239069)]
18. Oyibo K, Morita PP. Designing better exposure notification apps: the role of persuasive design. *JMIR Public Health Surveill* 2021 Nov 16;7(11):e28956 [FREE Full text] [doi: [10.2196/28956](https://doi.org/10.2196/28956)] [Medline: [34783673](https://pubmed.ncbi.nlm.nih.gov/34783673/)]
19. Sharma S, Singh G, Sharma R, Jones P, Kraus S, Dwivedi YK. Digital health innovation: exploring adoption of COVID-19 digital contact tracing apps. *IEEE Trans Eng Manage* (forthcoming) 2020 Sep 5:1-17 [FREE Full text] [doi: [10.1109/tem.2020.3019033](https://doi.org/10.1109/tem.2020.3019033)]
20. Walrave M, Waeterloos C, Ponnet K. Ready or not for contact tracing? Investigating the adoption intention of COVID-19 contact-tracing technology using an extended unified theory of acceptance and use of technology model. *Cyberpsychol Behav Soc Netw* 2021 Jun;24(6):377-383. [doi: [10.1089/cyber.2020.0483](https://doi.org/10.1089/cyber.2020.0483)] [Medline: [33017171](https://pubmed.ncbi.nlm.nih.gov/33017171/)]
21. Altmann S, Milsom L, Zillesen 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](https://doi.org/10.2196/19857)] [Medline: [32759102](https://pubmed.ncbi.nlm.nih.gov/32759102/)]



22. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. *Commun Assoc Inf Syst* 2009;24:485-500 [FREE Full text] [doi: [10.17705/1cais.02428](https://doi.org/10.17705/1cais.02428)]
23. Oyibo K. EMVE-DeCK: a theory-based framework for designing and tailoring persuasive technology. In: *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. 2021 Presented at: UMAP '21; June 21-25, 2021; Utrecht, The Netherlands p. 257-267 URL: <https://doi.org/10.1145/3450614.3464617> [doi: [10.1145/3450614.3464617](https://doi.org/10.1145/3450614.3464617)]
24. Wiafe I. A unified framework for analysing, designing and evaluating persuasive technologies. University of Reading. 2012 Sep. URL: [https://www.academia.edu/10613331/A\\_Framework\\_for\\_Analysing\\_Designing\\_and\\_Evaluating\\_Persuasive\\_Technologies](https://www.academia.edu/10613331/A_Framework_for_Analysing_Designing_and_Evaluating_Persuasive_Technologies) [accessed 2022-06-04]
25. Oyibo K, Yasunaga T, Morita PP. Designing exposure notification applications as persuasive technologies to improve uptake and effectiveness. In: *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*. 2021 Presented at: HFES '21; April 12-16, 2021; Virtual URL: <https://hfeshcs2021.conference-program.com/presentation/?id=INDLEC155&sess=sess102>
26. Federally-backed Covid Alert app now available in Ontario. CTV News. 2020. URL: <https://ottawa.ctvnews.ca/federally-backed-covidalert-app-now-available-in-ontario-1.5046667> [accessed 2022-01-31]
27. Oyibo K, Orji R, Vassileva J. Investigation of the persuasiveness of social influence in persuasive technology and the effect of age and gender. In: *Proceedings of the 2nd International Workshop on Personalization in Persuasive Technology*. 2017 Presented at: PPT '17; April 4, 2017; Amsterdam, The Netherlands p. 32-44. [doi: [10.1145/3099023.3099071](https://doi.org/10.1145/3099023.3099071)]
28. Oyibo K, Orji R, Vassileva J. The influence of culture in the effect of age and gender on social influence in persuasive technology. In: *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. 2017 Presented at: UMAP '17; July 9-12, 2017; Bratislava, Slovakia p. 47-52. [doi: [10.1145/3099023.3099071](https://doi.org/10.1145/3099023.3099071)]
29. Orji R, Lomotey R, Oyibo K, Orji F, Blustein J, Shahid S. Tracking feels oppressive and 'punishy': exploring the costs and benefits of self-monitoring for health and wellness. *Digit Health* 2018 Sep 3;4:2055207618797554 [FREE Full text] [doi: [10.1177/2055207618797554](https://doi.org/10.1177/2055207618797554)] [Medline: [30202544](https://pubmed.ncbi.nlm.nih.gov/30202544/)]
30. Oyibo K, Vassileva J. Persuasive features that drive the adoption of a fitness application and the moderating effect of age and gender. *Multimodal Technol Interact* 2020 May 11;4(2):17. [doi: [10.3390/mti4020017](https://doi.org/10.3390/mti4020017)]
31. Oyibo K. Investigating the key persuasive features for fitness app design and extending the persuasive system design model: a qualitative approach. *Proc Int Symp Human Factors Ergon Health Care* 2021 Jul 22;10(1):47-53. [doi: [10.1177/2327857921101022](https://doi.org/10.1177/2327857921101022)]
32. Orji R, Oyibo K, Lomotey RK, Orji FA. Socially-driven persuasive health intervention design: competition, social comparison, and cooperation. *Health Informatics J* 2019 Dec;25(4):1451-1484 [FREE Full text] [doi: [10.1177/1460458218766570](https://doi.org/10.1177/1460458218766570)] [Medline: [29801426](https://pubmed.ncbi.nlm.nih.gov/29801426/)]
33. Hair JF, Tatham RL, Anderson RE, Black W. *Multivariate Data Analysis*. 5th Edition. Hoboken, NJ, USA: Prentice Hall; 1998.
34. Oyibo K, Morita PP. COVID alert: factors influencing the adoption of exposure notification apps among Canadian residents. *Front Digit Health* 2022 Mar 11;4:842661 [FREE Full text] [doi: [10.3389/fdgth.2022.842661](https://doi.org/10.3389/fdgth.2022.842661)] [Medline: [35360366](https://pubmed.ncbi.nlm.nih.gov/35360366/)]
35. Oyibo K, Sahu KS, Oetomo A, Morita PP. Factors influencing the adoption of contact tracing applications: protocol for a systematic review. *JMIR Res Protoc* 2021 Jun 01;10(6):e28961 [FREE Full text] [doi: [10.2196/28961](https://doi.org/10.2196/28961)] [Medline: [33974551](https://pubmed.ncbi.nlm.nih.gov/33974551/)]
36. Hernández-Quevedo C, Scarpetti G, Webb E, Shuftan N, Williams GA, Birk HO, et al. Effective contact tracing and the role of apps: lessons from Europe. *Eurohealth* 2020;26(2):40-44 [FREE Full text]
37. Dowthwaite L, Fischer J, Perez Vallejos E, Portillo V, Nichele E, Goulden M, et al. Public adoption of and trust in the NHS COVID-19 contact tracing app in the United Kingdom: quantitative online survey study. *J Med Internet Res* 2021 Sep 17;23(9):e29085 [FREE Full text] [doi: [10.2196/29085](https://doi.org/10.2196/29085)] [Medline: [34406960](https://pubmed.ncbi.nlm.nih.gov/34406960/)]
38. Oyibo K, Sahu K, Oetomo A, Morita PP. Factors influencing the adoption of contact tracing applications: systematic review and recommendations. *Front Digit Health* 2022 May 3;4:862466 [FREE Full text] [doi: [10.3389/fdgth.2022.862466](https://doi.org/10.3389/fdgth.2022.862466)] [Medline: [35592459](https://pubmed.ncbi.nlm.nih.gov/35592459/)]
39. What is the Definition of Online Privacy? Winston & Strawn LLP. 2020. URL: <https://www.winston.com/en/legal-glossary/online-privacy.html> [accessed 2021-11-03]
40. Wogalter MS, Mayhorn CB. Trusting the Internet: cues affecting perceived credibility. *Int J Technol Human Interact* 2008;4(1):75-93 [FREE Full text] [doi: [10.4018/jthi.2008010105](https://doi.org/10.4018/jthi.2008010105)]
41. Nemeč Zlatolas L, Welzer T, Hölbl M, Heričko M, Kamišalić A. A model of perception of privacy, trust, and self-disclosure on online social networks. *Entropy (Basel)* 2019 Aug 07;21(8):772 [FREE Full text] [doi: [10.3390/e21080772](https://doi.org/10.3390/e21080772)] [Medline: [33267485](https://pubmed.ncbi.nlm.nih.gov/33267485/)]
42. Taddei S, Contena B. Privacy, trust and control: which relationships with online self-disclosure? *Comput Human Behav* 2013 May;29(3):821-826 [FREE Full text] [doi: [10.1016/j.chb.2012.11.022](https://doi.org/10.1016/j.chb.2012.11.022)]
43. Metzger MJ. Privacy, trust, and disclosure: exploring barriers to electronic commerce. *J Comput Mediat Commun* 2004 Jul;9(4):JCMC942. [doi: [10.1111/j.1083-6101.2004.tb00292.x](https://doi.org/10.1111/j.1083-6101.2004.tb00292.x)]
44. Riquelme IP, Román S. Is the influence of privacy and security on online trust the same for all type of consumers? *Electron Markets* 2014 Jan 22;24(2):135-149 [FREE Full text] [doi: [10.1007/s12525-013-0145-3](https://doi.org/10.1007/s12525-013-0145-3)]

45. Choon Ling K, Bin Daud D, Hoi Piew T, Keoy KH, Hassan P. Perceived risk, perceived technology, online trust for the online purchase intention in Malaysia. *Int J Bus Manag* 2011 Jun 01;6(6):167 [FREE Full text] [doi: [10.5539/ijbm.v6n6p167](https://doi.org/10.5539/ijbm.v6n6p167)]
46. Kaspar K. Motivations for social distancing and app use as complementary measures to combat the COVID-19 pandemic: quantitative survey study. *J Med Internet Res* 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: [10.2196/21613](https://doi.org/10.2196/21613)] [Medline: [32759100](https://pubmed.ncbi.nlm.nih.gov/32759100/)]
47. Velicia-Martin F, Cabrera-Sanchez JP, Gil-Cordero E, Palos-Sanchez PR. Researching COVID-19 tracing app acceptance: incorporating theory from the technological acceptance model. *PeerJ Comput Sci* 2021 Jan 4;7:e316 [FREE Full text] [doi: [10.7717/peerj-cs.316](https://doi.org/10.7717/peerj-cs.316)] [Medline: [33816983](https://pubmed.ncbi.nlm.nih.gov/33816983/)]
48. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mülken M. COVID-19 contact tracing apps: predicted uptake in the Netherlands based on a discrete choice experiment. *JMIR Mhealth Uhealth* 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: [10.2196/20741](https://doi.org/10.2196/20741)] [Medline: [32795998](https://pubmed.ncbi.nlm.nih.gov/32795998/)]
49. Thomas R, Michaleff ZA, Greenwood H, Abukmail E, Glasziou P. Concerns and misconceptions about the Australian government's COVIDSafe app: cross-sectional survey study. *JMIR Public Health Surveill* 2020 Nov 04;6(4):e23081 [FREE Full text] [doi: [10.2196/23081](https://doi.org/10.2196/23081)] [Medline: [33048826](https://pubmed.ncbi.nlm.nih.gov/33048826/)]
50. Timberg C, Harwell D, Safarpour A. Most Americans are not willing or able to use an app tracking coronavirus infections. That's a problem for Big Tech's plan to slow the pandemic. *The Washington Post*. 2020 Apr 29. URL: <https://www.washingtonpost.com/technology/2020/04/29/most-americans-are-not-willing-or-able-use-an-app-tracking-coronavirus-infections-thats-problem-big-techs-plan-slow-pandemic/> [accessed 2022-06-04]
51. van Loon MH. Self-assessment and self-reflection to measure and improve self-regulated learning in the workplace. In: McGrath S, Mulder M, Papier J, Suart R, editors. *Handbook of Vocational Education and Training: Developments in the Changing World of Work*. Cham, Switzerland: Springer; Jun 12, 2018:1-20.
52. McLeod S. Albert Bandura's Social Learning Theory. *SimplyPsychology*. 2016. URL: <https://www.simplypsychology.org/bandura.html> [accessed 2021-12-03]
53. Oyibo K. Designing Culture-Tailored Persuasive Technology to Promote Physical Activity. University of Saskatchewan. 2020. URL: <https://harvest.usask.ca/handle/10388/12943> [accessed 2022-06-04]
54. Baumeister RF, Gailliot M, DeWall CN, Oaten M. Self-regulation and personality: how interventions increase regulatory success, and how depletion moderates the effects of traits on behavior. *J Pers* 2006 Dec;74(6):1773-1801. [doi: [10.1111/j.1467-6494.2006.00428.x](https://doi.org/10.1111/j.1467-6494.2006.00428.x)] [Medline: [17083666](https://pubmed.ncbi.nlm.nih.gov/17083666/)]
55. Bandura A. Social Learning Theory. *MarcR Career Professionals*. URL: <https://marcr.net/marcr-for-career-professionals/career-theory/career-theories-and-theorists/social-learning-theory-bandura/> [accessed 2021-12-03]
56. Lyons SD, Berge ZL. Social Learning Theory. In: Seel NM, editor. *Encyclopedia of the Sciences of Learning*. Boston, MA, USA: Springer; 2012:1-6.
57. Oyibo K, Vassileva J. HOMEX: persuasive technology acceptance model and the moderating effect of culture. *Front Comput Sci* 2020 Mar 25;2:10 [FREE Full text] [doi: [10.3389/fcomp.2020.00010](https://doi.org/10.3389/fcomp.2020.00010)]
58. Oyibo K, Vassileva J. Relationship between perceived UX design attributes and persuasive features: a case study of fitness app. *Information* 2021 Sep 07;12(9):365 [FREE Full text] [doi: [10.3390/info12090365](https://doi.org/10.3390/info12090365)]
59. Chayinska M, Minescu A, McGarty C. 'The More We Stand For - The More We Fight For': compatibility and legitimacy in the effects of multiple social identities. *Front Psychol* 2017 Apr 26;8:642 [FREE Full text] [doi: [10.3389/fpsyg.2017.00642](https://doi.org/10.3389/fpsyg.2017.00642)] [Medline: [28491046](https://pubmed.ncbi.nlm.nih.gov/28491046/)]
60. Hubert M, Blut M, Brock C, Zhang RW, Koch V, Riedl R. The influence of acceptance and adoption drivers on smart home usage. *Eur J Mark* 2019 Jun 10;53(6):1073-1098. [doi: [10.1108/EJM-12-2016-0794](https://doi.org/10.1108/EJM-12-2016-0794)]
61. Taylor S, Todd PA. Understanding information technology usage: a test of competing models. *Inf Syst Res* 1995 Jun;6(2):144-176. [doi: [10.1287/isre.6.2.144](https://doi.org/10.1287/isre.6.2.144)]
62. Introduction to persuasion. YouTube. 2018 Nov 4. URL: <https://www.youtube.com/watch?v=iufD8CeQpAo> [accessed 2021-02-19]
63. Oyibo K, Afaji I, Orji R, Olabenjo B, Vassileva J. The interplay between classical aesthetics, expressive aesthetics and persuasiveness in behavior modeling. In: *Proceedings of the 32nd International BCS Human Computer Interaction Conference*. 2018 Presented at: HCI '18; July 4-6, 2018; Belfast, UK p. 1-10. [doi: [10.14236/ewic/hci2018.15](https://doi.org/10.14236/ewic/hci2018.15)]
64. Wang H, Lee K. Getting in the flow together: the role of social presence, perceived enjoyment and concentration on sustainable use intention of mobile social network game. *Sustainability* 2020 Aug 24;12(17):6853 [FREE Full text] [doi: [10.3390/su12176853](https://doi.org/10.3390/su12176853)]
65. Lehto T, Oinas-Kukkonen H, Drozd F. Factors affecting perceived persuasiveness of a behavior change support system. In: *Proceedings of the 33rd International Conference on Information Systems*. 2012 Presented at: ICIS '12; December 16-19, 2012; Orlando, FL, USA p. 1-15.
66. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspect Psychol Sci* 2011 Jan;6(1):3-5. [doi: [10.1177/1745691610393980](https://doi.org/10.1177/1745691610393980)] [Medline: [26162106](https://pubmed.ncbi.nlm.nih.gov/26162106/)]
67. Welcome to the Amazon Mechanical Turk Requester User Interface Guide. Amazon Web Services. URL: <https://docs.aws.amazon.com/AWSMechTurk/latest/RequesterUI/Introduction.html> [accessed 2021-11-03]

68. Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restor Dent Endod* 2017 May;42(2):152-155 [FREE Full text] [doi: [10.5395/rde.2017.42.2.152](https://doi.org/10.5395/rde.2017.42.2.152)] [Medline: [28503482](https://pubmed.ncbi.nlm.nih.gov/28503482/)]
69. Brant R. Inference for Means: Comparing Two Independent Samples. University of British Columbia. 2012. URL: <http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html> [accessed 2021-03-17]
70. Orji R. Persuasion and culture: individualism–collectivism and susceptibility to influence strategies. In: Proceedings of the 2016 Personalization in Persuasive Technology Workshop. 2016 Presented at: PPT '16; April 5, 2016; Salzburg, Austria.
71. Al-Jabri IM. The perceptions of adopters and non-adopters of cloud computing: application of technology-organization-environment framework. In: Proceedings of the 14th International Conference of Electronic Business. 2014 Presented at: ICEB '14; December 8-12, 2014; Taipei, Taiwan.
72. Emani S, Peters E, Desai S, Karson AS, Lipsitz SR, LaRocca R, et al. Who adopts a patient portal?: an application of the diffusion of innovation model. *J Innov Health Inform* 2018 Oct 25;25(3):149-157 [FREE Full text] [doi: [10.14236/jhi.v25i3.991](https://doi.org/10.14236/jhi.v25i3.991)] [Medline: [30398458](https://pubmed.ncbi.nlm.nih.gov/30398458/)]
73. Dickerson MD, Gentry JW. Characteristics of adopters and non-adopters of home computers. *J Consum Res* 1983 Sep;10(2):225-235 [FREE Full text] [doi: [10.1086/208961](https://doi.org/10.1086/208961)]
74. Sanchez G. PLS Path Modeling with R. Berkley. 2013. URL: [https://www.gastonsanchez.com/PLS\\_Path\\_Modeling\\_with\\_R.pdf](https://www.gastonsanchez.com/PLS_Path_Modeling_with_R.pdf) [accessed 2022-06-04]
75. Hair Jr JF, Hult GT, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2nd edition. Washington, DC, USA: Sage Publications; 2016.
76. van den Berg RG. Effect Size – A Quick Guide. SPSS Tutorials. URL: <https://www.spss-tutorials.com/effect-size/> [accessed 2021-09-29]
77. Mangiafico SS. An R Companion for the Handbook of Biological Statistics. Version 1.3.2. R Companion. 2015. URL: [https://rcompanion.org/rcompanion/a\\_02.html](https://rcompanion.org/rcompanion/a_02.html) [accessed 2022-06-04]
78. Kelley K, Stanley D. Package 'effectsize'. The Comprehensive R Archive Network. 2021. URL: <https://cran.r-project.org/web/packages/effectsize/effectsize.pdf> [accessed 2021-09-23]
79. Kotrlik J, Williams H, Jabor K. Reporting and interpreting effect size in quantitative agricultural education research. *J Agric Educ* 2011 Mar 01;52(1):132-142 [FREE Full text] [doi: [10.5032/jae.2011.01132](https://doi.org/10.5032/jae.2011.01132)]
80. Hussain S, Fangwei Z, Siddiqi AF, Ali Z, Shabbir MS. Structural Equation Model for evaluating factors affecting quality of social infrastructure projects. *Sustainability* 2018 May 03;10(5):1415 [FREE Full text] [doi: [10.3390/su10051415](https://doi.org/10.3390/su10051415)]
81. Abrahams N, Cwalina C, Evans M, Flockhart F, Gamvros A, Lennon J. Contact tracing apps in Australia: A new world for data privacy. Norton Rose Fulbright. 2021. URL: <https://www.nortonrosefulbright.com/-/media/files/nrf/nrfweb/contact-tracing/australia-contact-tracing.pdf?revision=9f35a88a-4124-4c48-b38f-68e86a187050&la=en> [accessed 2021-01-31]
82. Abuhammad S, Khabour OF, Alzoubi KH. Covid-19 contact-tracing technology: acceptability and ethical issues of use. *Patient Prefer Adherence* 2020 Sep 18;14:1639-1647 [FREE Full text] [doi: [10.2147/PPA.S276183](https://doi.org/10.2147/PPA.S276183)] [Medline: [32982188](https://pubmed.ncbi.nlm.nih.gov/32982188/)]
83. Saint-Arnaud P. Ottawa spent nearly \$20 million on COVID-19 tracking app -- with inconclusive results. CTV News. 2021 Jul 6. URL: <https://www.ctvnews.ca/politics/ottawa-spent-nearly-20-million-on-covid-19-tracking-app-with-inconclusive-results-1.5497296> [accessed 2022-06-02]

## Abbreviations

**GOF:** goodness of fit

**PSD:** persuasive system design

**RM-ANOVA:** repeated-measure ANOVA

**TAM:** Technology Acceptance Model

*Edited by A Mavragani; submitted 11.10.21; peer-reviewed by B Marcolin, V Mylonopoulou; comments to author 01.11.21; revised version received 08.12.21; accepted 29.04.22; published 06.09.22.*

*Please cite as:*

Oyibo K, Morita PP

*The Effect of Persuasive Design on the Adoption of Exposure Notification Apps: Quantitative Study Based on COVID Alert*

*JMIR Form Res* 2022;6(9):e34212

URL: <https://formative.jmir.org/2022/9/e34212>

doi:[10.2196/34212](https://doi.org/10.2196/34212)

PMID:[35580138](https://pubmed.ncbi.nlm.nih.gov/35580138/)

©Kiemute Oyibo, Plinio Pelegrini Morita. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 06.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License

(<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Circulating Illness and Changes in Thermometer Use Behavior: Series of Cross-sectional Analyses

Jack Seifarth<sup>1,2</sup>, MPH; Megan Pinaire<sup>2,3</sup>, MPH; John Zicker<sup>2</sup>, MSEE; Inder Singh<sup>2</sup>, MBA, MPP, MS; Danielle Bloch<sup>2</sup>, MPH

<sup>1</sup>Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, United States

<sup>2</sup>Kinsa Inc, San Francisco, CA, United States

<sup>3</sup>Department of Biostatistics, Yale School of Public Health, New Haven, CT, United States

**Corresponding Author:**

Jack Seifarth, MPH

Department of Epidemiology

Mailman School of Public Health

Columbia University

722 West 168th St

New York, NY, 10032

United States

Phone: 1 717 887 2685

Email: [jack.seifarth@kinsahealth.com](mailto:jack.seifarth@kinsahealth.com)

## Abstract

**Background:** Temperature-taking behaviors vary with levels of circulating infectious illness; however, little is known about how these behaviors differ by demographic characteristics. Populations with higher perceived risks of illness are more likely to adopt protective health behaviors.

**Objective:** We investigated differences in temperature-taking frequency and the proportion of readings that were feverish among demographic groups (age, gender, urban/rural status) over influenza offseason; influenza season; and waves 1, 2, and 3 of the COVID-19 pandemic.

**Methods:** Using data from smart thermometers collected from May 1, 2019, to February 28, 2021, across the United States, we calculated the frequency of temperature-taking and the proportion of temperature readings that were feverish. Mixed-effects negative binomial and mixed-effects logistic regression analyses were performed to identify demographic characteristics associated with temperature-taking frequency and the proportion of feverish readings, respectively. Separate models were fit over five study periods: influenza offseason (n=122,480), influenza season (n=174,191), wave 1 of COVID-19 (n=350,385), wave 2 (n=366,489), and wave 3 (n=391,578).

**Results:** Both temperature-taking frequency and the proportion of feverish readings differed by study period (ANOVA  $P < .001$ ) and were the highest during influenza season. During all periods, children aged 2-5 years and 6-11 years had significantly higher frequencies of temperature-taking than users aged 19-30 years, and children had the highest proportion of feverish readings of all age groups, after adjusting for covariates. During wave 1 of COVID-19, users over the age of 60 years had 1.79 times (95% CI 1.76-1.83) the rate of temperature-taking as users aged 19-30 years and 74% lower odds (95% CI 72%-75%) of a reading being feverish. Across all periods, men had significantly lower temperature-taking frequency and significantly higher odds of having a feverish reading compared to women. Users living in urban areas had significantly higher frequencies of temperature-taking than rural users during all periods, except wave 2 of COVID-19, and urban users had higher odds of a reading being feverish in all study periods except wave 1 of COVID-19.

**Conclusions:** Temperature-taking behavior and the proportion of readings that were feverish are associated with both population disease levels and individual demographic characteristics. Differences in the health behavior of temperature-taking may reflect changes in both perceived and actual illness risk. Specifically, older adults may have experienced an increase in perceived risk during the first three waves of COVID-19, leading to increased rates of temperature monitoring, even when their odds of fever were lower than those of younger adults. Men's perceived risk of circulating infectious illnesses such as influenza and COVID-19 may be lower than that of women, since men took their temperature less frequently and each temperature had a higher odds of being feverish across all study periods. Infectious disease surveillance should recognize and incorporate how behavior impacts illness monitoring and testing.

**KEYWORDS**

thermometer; health behavior; influenza; COVID-19; fever; surveillance; perceived risk; percent positivity; smart technology; smart thermometer; population demography; older adult; elderly population; health monitoring

## Introduction

At-home health monitoring behaviors have the potential to greatly impact health outcomes. However, health behaviors differ by demographic and social determinants, including poverty, gender, and neighborhood social and physical characteristics [1]. For example, women and older individuals are more likely to report practicing protective health behaviors [2] such as taking COVID-19 precautions [3]. Individuals also alter behaviors in response to circulating illness levels and associated health recommendations or policies; this was observed with both increases in mask wearing and social distancing after the H1N1 influenza outbreak [4], and increases in handwashing and social distancing during the COVID-19 pandemic [5].

The Health Action Process Approach (HAPA) framework for health behavior states that perceived risk, particularly of severe health outcomes, can motivate health behavior change [6,7]. Additionally, changes to perceived risk based on underlying conditions, attention to media coverage, or knowledge of disease can impact health behaviors [8,9]. A study of behavior among individuals in the United States during spring of 2020 guided by HAPA found that risk perception and self-efficacy were both predictors of social distancing [10]. A systematic review of nonpharmaceutical interventions prior to the COVID-19 pandemic found that individuals adopt behaviors partially based on their perceived vulnerability of respiratory illness [11].

Most respiratory illness-related health behavior studies are cross-sectional surveys relying on self-reported behavior during a pandemic or influenza season [12-14]. Temperature-taking using a smart thermometer is a timely and sensitive surveillance measure that circumvents the issue of self-reporting. Smart thermometers record body temperatures and aggregate the anonymized, deidentified readings along with basic demographic information [15]. Based on a user's temperature, symptoms, and age, they receive illness guidance through an associated app. Aggregated user temperature data are closely correlated with traditional influenza surveillance methods [16-18]. Unlike COVID-19 prevention behaviors such as mask wearing and social distancing, temperature-taking is not typically performed in public settings. Therefore, temperature-taking may be less influenced by social pressures and could therefore better approximate individual perceptions of risk. Previous studies using smart thermometers have shown that the number of fevers and the total number of thermometer readings correlate with influenza-like illness at the national and regional levels during both influenza season [16,17] and offseason [16], and with influenza test positivity at the regional level [17]. Kinsa fever data were also found to be correlated with confirmed cases during the first wave of COVID-19 [19].

The demographic and social determinants of the health behavior of temperature-taking have yet to be examined. We predicted that demographic groups with greater perceived risks of influenza or COVID-19 will take their temperature more often. Using data from smart thermometer users collected between May 1, 2019, and February 28, 2021, we aimed to understand how both demographic characteristics and external circulating illness were associated with temperature-taking frequency and the percent of readings with a fever. We use temperature-taking frequency as a proxy for an individual's perceived risk of illness and percent feverish readings to determine if there was febrile illness in the presence of that concern. We sought to explore how both social determinants of health and external influenza and COVID-19 levels impacted perceived and actual risks of febrile illness.

## Methods

### Data Collection

Kinsa smart thermometers record and store body temperatures using a smartphone app. Most users purchase their thermometer through major retailers. Kinsa thermometers are also distributed free of charge for families in Title 1 elementary schools through a program called FLUency [20]. Title 1 programs provide federal funding to schools with high numbers or percentages of children from low-income families [21]. FLUency school nurses can use the program to communicate with families about current illness in the school or grade. When any user takes a temperature, the reading and timestamp are recorded along with deidentified, user-entered demographic information, including age and gender. Readings are geocoded using GPS coordinates or the IP address of the connected device. Users can assign temperature readings to different profiles within their account, allowing for differentiation among readings from multiple users in the same household.

### Study Population

Individuals who recorded at least one temperature reading with a Kinsa thermometer in the United States from May 1, 2019, to February 28, 2021, were included in this analysis. Study periods were defined based on trends of seasonal influenza and COVID-19: influenza offseason (May 1, 2019, to October 31, 2019), influenza season (November 1, 2019, to February 2, 2020), wave 1 of COVID-19 (February 3, 2020, to May 31, 2020), wave 2 of COVID-19 (June 1, 2020, to October 31, 2020), and wave 3 of COVID-19 (November 1, 2020, to February 28, 2021). The 2019-2020 influenza season was considered moderately severe and was dominated by A(H1N1)pdm09 viruses [22]. The influenza season ended in early 2020, likely due to COVID-19 lockdowns and precautions [22].

Users were only included in a study period if they recorded at least one temperature reading within that period. Therefore, the

same user would not be included in all study periods if they did not record a reading in each separate period.

We performed our analyses on users who had no missing demographic or geographic information. Standardized mean differences showed that the full population had a similar number of readings as the complete case population.

## Measures

The two outcomes assessed were: (1) the number of readings per user and (2) the proportion of readings with a fever. The number of readings provides a measure of how often a user is potentially concerned about a possible fever and the proportion of readings with a fever provides a measure of how many readings were taken because of a true fever. To define the number of readings per user, we counted each reading a user took during a given period and then adjusted for varying amounts of follow-up time. The number of days a user was active during a study period was calculated from device activation through to the end of that period. If the activation date was before the start of the study period, the user was considered to be active for the entire period. Because users could take multiple readings on the same day, a sensitivity analysis was performed that examined the number of distinct days with at least one temperature reading to circumvent intraday variability in reading behavior. The second outcome, proportion of readings with a fever, was defined as the number of readings  $>37.8^{\circ}\text{C}$  divided by the total number of readings for a user during a period.

Age and gender were self-reported and defined at the first reading during the period. Age was categorized into 0-1 years, 2-5 years, 6-11 years, 12-18 years, 19-30 years, 31-60 years, and 61+ years. We assumed that an adult in the household was driving temperature-taking among individuals 18 years and under. Any user associated with a device that was distributed through the school program was categorized as a FLUency user.

Household composition was derived from the ages of registered users associated with one thermometer: if all users were  $<18$  years old, the household was considered “child only”; if all users were aged 18 years and older, the household was considered “adult only”; and if there were users both under 18 years and 18 years and older, the household was categorized as “multigeneration.” “Child-only” households reflect devices where a parent/guardian has not created a profile for themselves but has created profiles for their children.

Neighborhood poverty, US region, and urban/rural designation were determined based on the location with a majority of a user’s readings. Census tract-level poverty was obtained from the 2015-2019 American Community Survey and defined as the percent of residents living below 100% of the federal poverty level [23]. Poverty was then classified into four categories: 0 to  $<10\%$ ,  $10\%$  to  $<20\%$ ,  $20\%$  to  $<30\%$ , and  $\geq 30\%$ . Region was defined using the 10 regions created by the Centers for Disease Control and Prevention National Center for Chronic Disease Prevention and Health Promotion [24]. A user was categorized as living in an urban area if 50% or more of the land area of their census tract (based on the 2010 US Census) was classified as urban [25].

## Statistical Analysis

We assessed unadjusted differences in both temperature-taking frequency and the proportion of readings with a fever across periods using ANOVA. Differences in the frequency of the categorical variables across periods were assessed by  $\chi^2$  tests.

We used mixed-effects negative binomial models to examine the relationships between explanatory variables and the overdispersed outcome of frequency of readings per month. A separate model was fit for each of the five study periods. Since users were nested within devices, device was treated as a random effect. All other variables of interest were treated as fixed effects. The outcome of the number of readings was offset by the number of days a user was active during the period to obtain a frequency of readings over time. This offset calculated a more conservative estimate of variance. We checked for collinearity among predictor variables by requiring their generalized variance inflation factors to be less than 5. Age group, gender, urban/rural status, census tract poverty group, household composition, region, and FLUency participation were included in the final model. A predictor was considered significant if the 95% CI of its incidence rate ratio (IRR) did not contain the null value of 1. The same methods were applied to the outcome of the distinct number of days with a reading in the sensitivity analysis.

We used mixed-effects logistic (binomial) regression to examine the adjusted relationships between our explanatory variables and the proportion of readings with a fever. A separate model was fit for each of the five study periods with the outcome of number of fevers and number of nonfevers recorded per user. Device was added as a random intercept with all other predictors treated as fixed effects. A predictor was considered significant if the 95% CI of its odds ratio (OR) did not contain the null value of 1.

Statistical analyses were performed in R 4.1.0 (R Core Team, Vienna, Austria 2021) using the glmmTMB package [26]. Reported regression outputs and ratios are adjusted for all predictors included in the model.

## Ethical Considerations

Upon downloading the app, users were asked to acknowledge and consent to data collection practices outlined in the Kinsa Privacy Policy [15]. Users must expressly consent to sharing geolocation data. All personally identifiable information was collected and maintained in compliance with state and federal confidentiality guidelines. This study was approved by an external institutional review board, Advarra Inc (Pro00065469).

## Results

### Descriptive Statistics

There were 122,480 users with full demographic information in influenza offseason 2019, 174,191 users in influenza season 2019-20, 350,385 users during wave 1 of COVID-19, 366,489 users during wave 2, and 391,578 users during wave 3 (Table 1). The combined study population had a median age of 26 years (IQR 6-43), with 57.3% women, and primarily resided in urban (77.8%) and 0-10% poverty tracts (56.0%); 12.9% of users

obtained their thermometer through the FLUency program. The study populations changed significantly over time (Table 1). Notably, the median age of the study population increased from 7 years during the influenza offseason to 30 years in wave 1 of COVID-19.

Temperature-taking frequency differed significantly by study period (Figure 1;  $F=141.2$ ,  $P<.001$ ). The median frequency of readings was lower during influenza offseason compared to that during influenza season (1.22 vs 2.06 readings per month).

During the COVID-19 pandemic, the median reading frequency was the highest during wave 1 (1.79 readings per user per month) and decreased slightly during waves 2 and 3 (1.76 and 1.23 readings per month, respectively).

The proportion of readings with a fever also differed significantly by study period (Figure 1;  $P<.001$ ). The mean percent of readings that were feverish was the highest during the influenza season (21.2%) and was the lowest during wave 2 of COVID-19 (5.6%).



**Table 1.** Demographic characteristics of the study population with temperature readings, stratified by study period, May 2019 to February 2021.

Characteristics	Offseason (May 1, 2019, to October 31, 2019)	Flu season (November 1, 2019, to February 2, 2020)	Wave 1 of COVID-19 (February 3, 2020, to May 31, 2020)	Wave 2 of COVID-19 (June 1, 2020, to October 31, 2020)	Wave 3 of COVID-19 (November 1, 2020, to February 28, 2021)	P value
Users, n	122,480	174,191	350,385	366,489	391,578	N/A <sup>a</sup>
Devices, n	83,628	113,433	219,056	239,010	255,483	N/A
Temperature readings per user, median (IQR)	3 (1-8)	3 (1-9)	4 (1-13)	4 (1-12)	3 (1-9)	N/A
Follow-up time (days), median (IQR)	130 (44-184)	69 (32-94)	79 (63-119)	109 (49-153)	105 (72-120)	N/A
Temperature readings/month, median (IQR)	1.22 (0.41-3.97)	2.06 (0.72-6.52)	1.79 (0.65-5.75)	1.76 (0.60-5.81)	1.23 (0.51-3.56)	<.001 <sup>b</sup>
Days with a temperature reading, median (IQR)	1 (1-3)	1 (1-3)	2 (1-4)	2 (1-5)	2 (1-4)	N/A
Days with a temperature reading/month, median (IQR)	0.53 (0.25-1.45)	0.97 (0.46-2.17)	0.81 (0.42-2.10)	0.87 (0.24-1.52)	0.72 (0.34-1.65)	<.001 <sup>b</sup>
Readings with a fever (%), median (IQR)	0 (0-0.30)	0 (0-0.39)	0 (0-0)	0 (0-0)	0 (0-0)	<.001 <sup>b</sup>
Age (years), median (IQR)	7 (2-30)	8 (3-30)	30 (8-48)	28 (9-46)	29 (9-46)	N/A
<b>Age group, n (%)</b>						<.001 <sup>c</sup>
0-1	20,146 (16.4)	17,898 (10.3)	18,914 (5.4)	20,325 (5.5)	18,329 (4.7)	
2-5	27,520 (22.5)	35,372 (20.3)	38,276 (10.9)	34,784 (9.5)	36,221 (9.3)	
6-11	22,387 (18.3)	45,924 (26.4)	46,554 (13.3)	47,890 (13.1)	57,647 (14.7)	
12-18	7127 (5.8)	14,751 (8.5)	23,659 (6.8)	27,857 (7.6)	28,559 (7.3)	
19-30	13,305 (10.9)	16,312 (9.4)	42,521 (12.1)	58,690 (16.0)	59,279 (15.1)	
31-60	28,935 (23.6)	40,624 (23.3)	128,326 (36.6)	130,512 (35.6)	139,048 (35.5)	
61+	3060 (2.5)	3310 (1.9)	52,135 (14.9)	46,431 (12.7)	52,495 (13.4)	
<b>Gender, n (%)</b>						<.001 <sup>c</sup>
Women	69,577 (56.8)	100,016 (57.4)	198,125 (56.5)	210,423 (57.4)	226,531 (57.9)	
Men	52,903 (43.2)	74,175 (42.6)	152,260 (43.5)	156,066 (42.6)	165,047 (42.1)	
<b>Poverty group<sup>d</sup>, n (%)</b>						<.001 <sup>c</sup>
0 to <10%	69,099 (56.4)	92,356 (53.0)	207,678 (59.3)	208,457 (56.9)	208,693 (53.3)	
10 to <20%	36,370 (29.7)	54,457 (31.3)	96,911 (27.7)	100,535 (27.4)	116,984 (29.9)	
20 to <30%	12,046 (9.8)	19,192 (11.0)	31,512 (9.0)	37,045 (10.1)	42,959 (11.0)	
≥30%	4965 (4.1)	8186 (4.7)	14,284 (4.1)	20,452 (5.6)	22,942 (5.9)	
<b>Density<sup>e</sup>, n (%)</b>						<.001 <sup>c</sup>
Urban tract (%)	96,212 (78.6)	127,192 (73.0)	282,673 (80.7)	291,516 (79.5)	295,074 (75.4)	
Rural tract (%)	26,268 (21.4)	46,999 (27.0)	67,712 (19.3)	74,973 (20.5)	96,504 (24.6)	
<b>FLUency group<sup>f</sup>, n (%)</b>						<.001 <sup>c</sup>
FLUency user	7673 (6.3)	46,809 (26.9)	31,695 (9.0)	33,402 (9.1)	66,930 (17.1)	
Non-FLUency user	114,807 (93.7)	127,382 (73.1)	318,690 (91.0)	333,087 (90.9)	324,648 (82.9)	
<b>Household composition<sup>g</sup>, n (%)</b>						<.001 <sup>c</sup>
Adult-only house	24,170 (19.7)	31,706 (18.2)	167,023 (47.7)	180,067 (49.1)	192,482 (49.2)	
Child-only house	53,461 (43.6)	78,542 (45.1)	70,015 (20.0)	71,099 (19.4)	77,866 (19.9)	

Characteristics	Offseason (May 1, 2019, to October 31, 2019)	Flu season (November 1, 2019, to February 2, 2020)	Wave 1 of COVID-19 (February 3, 2020, to May 31, 2020)	Wave 2 of COVID-19 (June 1, 2020, to October 31, 2020)	Wave 3 of COVID-19 (November 1, 2020, to February 28, 2021)	P value
Multigenerational house (%)	44,849 (46.6)	63,943 (36.7)	113,347 (32.3)	115,323 (31.5)	121,230 (31.0)	
<b>Region<sup>h</sup> (comprising state abbreviations), n (%)</b>						<.001 <sup>c</sup>
1 (CT, ME, RI, MA, NH, NY, VT)	11,511 (9.4)	14,681 (8.4)	39,386 (11.2)	53,182 (14.5)	44,391 (11.3)	
2 (DC, MD, WV, DE, NJ, PA, VA)	15,474 (12.6)	21,808 (12.5)	44,449 (12.7)	40,462 (11.0)	63,308 (16.2)	
3 (GA, FL, NC, SC)	17,207 (14.0)	22,567 (13.0)	40,731 (11.6)	37,444 (10.2)	37,389 (9.5)	
4 (KY, TN, AL, MS)	5918 (4.8)	10,579 (6.1)	12,699 (3.6)	12,317 (3.4)	17,136 (4.4)	
5 (IL, WI, IN, MI, MN, OH)	19,509 (15.9)	28,607 (16.4)	57,017 (16.3)	61,453 (16.8)	61,028 (15.6)	
6 (OK, AR, LA, NM, TX)	17,547 (14.3)	28,915 (16.6)	39,731 (11.3)	38,069 (10.4)	45,840 (11.7)	
7 (NE, IA, KS, MO)	7053 (5.8)	11,578 (6.6)	16,867 (4.8)	18,975 (5.2)	18,531 (4.7)	
8 (MT, ND, WY, CO, SD, UT)	3345 (2.7)	4343 (2.5)	12,129 (3.5)	13,998 (3.8)	13,627 (3.5)	
9 (CA, NV, AZ, HI)	20,990 (17.1)	26,310 (15.1)	70,611 (20.2)	73,715 (20.1)	75,880 (19.4)	
10 (AK, ID, OR, WA)	3926 (3.2)	4803 (2.8)	16,765 (4.8)	16,874 (4.6)	14,448 (3.7)	

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

<sup>b</sup>Differences in continuous variables across study periods assessed via ANOVA.

<sup>c</sup>Differences in frequencies of categorical variables across study periods assessed via  $\chi^2$  test.

<sup>d</sup>Percentage of population living below the 100% federal poverty level at the census tract level, from the 2015-2019 American Community Survey.

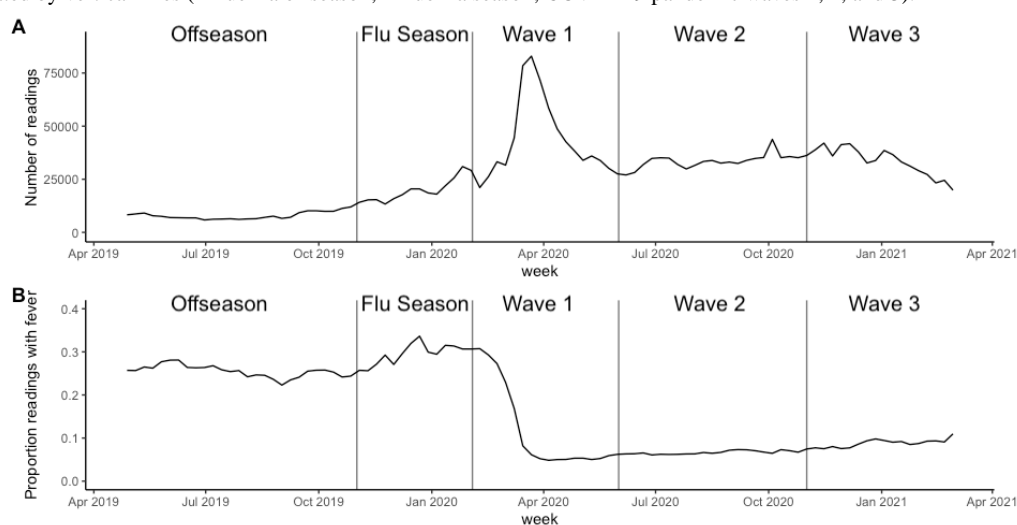
<sup>e</sup>Categorized as urban if the census tract was part of an urbanized area of 50,000 or more people based on the 2010 US Census.

<sup>f</sup>Received the thermometer through Kinsa's school distribution and engagement program, FLUency.

<sup>g</sup>Based on ages of profiles associated with the device, with child-only households representing devices where a parent has made profiles for their children but not themselves.

<sup>h</sup>Classified using the Centers for Disease Control and Prevention National Center For Chronic Disease Prevention and Health Promotion Regions.

**Figure 1.** The number of thermometer readings (A) and proportion of readings with a fever (B) aggregated by week, May 2019-February 2021. Study periods are separated by vertical lines (influenza offseason; influenza season; COVID-19 pandemic waves 1, 2, and 3).



### Temperature-Taking Frequency Regression

During all study periods, the age groups of 0-1, 2-5, and 6-11 years had significantly higher rates of temperature-taking than

those of users aged 19-30 years (Table 2). During influenza season, users aged 6-11 years had elevated rates of temperature-taking (IRR 2.25, 95% CI 2.18-2.31) and users aged over 60 years had suppressed rates of temperatures-taking

(IRR 0.94, 95% CI 0.89-1.00) compared to users aged 19-30 years. During wave 1 of the COVID-19 pandemic, these age patterns reversed, with users aged over 60 years taking temperatures at a significantly increased rate (IRR 1.79, 95% CI 1.76-1.83) compared to those of young adults. During COVID-19 waves 2 and 3, users aged over 60 years continued to have significantly higher frequencies of temperature-taking ([Table 2](#)).

Men had a significantly lower rate of temperature-taking compared to women during all periods ([Table 2](#)), and the

difference was the largest during influenza offseason (IRR 0.91, 95% CI 0.90-0.93). Users living in urban census tracts had an increased rate of temperature-taking compared to that of rural users during all periods except wave 2 of COVID-19, when urban users had 0.95 (95% CI 0.93-0.96) times the rate of temperature-taking compared to rural users. Similar trends for age, gender, and population density were observed in the sensitivity analysis using the outcome of distinct days with at least one reading ([Multimedia Appendix 1](#)).

**Table 2.** Characteristics associated with temperature-taking frequency from multivariable mixed-effects negative binomial regressions, May 2019-February 2021.

Variable	Incidence rate ratio (95% CI) <sup>a</sup>				
	Offseason 2019 (May 1, 2019, to October 31, 2019)	Flu season (Novem- ber 1, 2019, to February 2, 2020)	Wave 1 of COVID-19 (February 3, 2020, to May 31, 2020)	Wave 2 of COVID- 19 (June 1, 2020, to October 31, 2020)	Wave 3 of COVID-19 (November 1, 2020, to February 28, 2021)
<b>Age (ref: 19-30 years)</b>					
0-1 years	2.41 (2.33-2.50)	1.65 (1.59-1.70)	1.78 (1.73-1.83)	1.34 (1.30-1.37)	1.36 (1.32-1.39)
2-5 years	2.37 (2.29-2.45)	2.23 (2.16-2.29)	1.71 (1.67-1.74)	1.26 (1.23-1.29)	1.29 (1.26-1.31)
6-11 years	1.90 (1.84-1.97)	2.25 (2.18-2.31)	1.40 (1.37-1.43)	1.11 (1.09-1.14)	1.08 (1.06-1.10)
12-18 years	1.40 (1.34-1.47)	1.67 (1.61-1.72)	1.03 (1.01-1.06)	0.94 (0.92-0.96)	0.92 (0.90-0.94)
31-60 years	0.92 (0.90-0.95)	0.96 (0.93-0.99)	1.32 (1.30-1.34)	1.02 (1.01-1.04)	1.08 (1.07-1.10)
>60 years	1.05 (0.98-1.11)	0.94 (0.89-1.00)	1.79 (1.76-1.83)	1.28 (1.25-1.31)	1.28 (1.26-1.31)
Gender, men (ref: women)	0.91 (0.90-0.93)	0.94 (0.92-0.95)	0.93 (0.92-0.94)	0.96 (0.95-0.97)	0.95 (0.94-0.96)
Density, urban (ref: rural) <sup>b</sup>	1.06 (1.03-1.09)	1.12 (1.09-1.14)	1.08 (1.06-1.10)	0.95 (0.93-0.96)	1.03 (1.02-1.04)
<b>Poverty (ref: 0 to &lt;10%)<sup>c</sup></b>					
10% to <20%	0.98 (0.95-1.00)	0.98 (0.96-1.00)	1.03 (1.01-1.04)	1.02 (1.00-1.03)	1.00 (0.99-1.02)
20% to <30%	0.97 (0.94-1.01)	0.97 (0.94-1.00)	1.07 (1.05-1.09)	0.99 (0.97-1.01)	0.95 (0.93-0.97)
≥30%	0.98 (0.93-1.03)	0.95 (0.91-0.99)	1.09 (1.06-1.12)	0.97 (0.94-1.00)	0.93 (0.91-0.95)
FLUency user (ref: non-FLUen- cy) <sup>d</sup>	6.00 (5.73-6.29)	0.65 (0.63-0.66)	0.72 (0.70-0.73)	1.31 (1.28-1.34)	0.67 (0.65-0.68)
<b>Household composition (ref: adult-only)<sup>e</sup></b>					
Child-only	0.48 (0.46-0.49)	0.49 (0.48-0.51)	0.51 (0.50-0.52)	0.66 (0.64-0.67)	0.70 (0.69-0.71)
Multigenerational	0.58 (0.57-0.60)	0.53 (0.51-0.54)	0.50 (0.50-0.51)	0.56 (0.55-0.57)	0.67 (0.66-0.68)
<b>Region (ref: 1 [Northeast])<sup>f</sup></b>					
2 (DC, MD, WV, DE, NJ, PA, VA)	0.94 (0.90-0.98)	0.88 (0.84-0.91)	0.90 (0.88-0.92)	0.91 (0.89-0.93)	0.85 (0.84-0.87)
3 (GA, FL, NC, SC)	1.04 (1.00-1.09)	0.86 (0.83-0.90)	0.83 (0.81-0.85)	0.90 (0.88-0.92)	1.04 (1.02-1.07)
4 (KY, TN, AL, MS)	0.92 (0.87-0.98)	0.86 (0.82-0.90)	0.80 (0.77-0.83)	0.87 (0.84-0.90)	0.96 (0.93-0.99)
5 (IL, WI, IN, MI, MN, OH)	0.99 (0.95-1.03)	0.87 (0.84-0.91)	0.88 (0.86-0.90)	1.01 (0.99-1.03)	0.92 (0.90-0.94)
6 (OK, AR, LA, NM, TX)	0.96 (0.93-1.01)	0.80 (0.77-0.83)	0.82 (0.80-0.84)	0.79 (0.77-0.81)	0.98 (0.96-1.00)
7 (NE, IA, KS, MO)	0.88 (0.84-0.93)	0.77 (0.73-0.80)	0.82 (0.79-0.85)	0.93 (0.90-0.96)	0.84 (0.82-0.87)
8 (MT, ND, WY, CO, SD, UT)	0.90 (0.84-0.97)	0.91 (0.85-0.96)	0.89 (0.85-0.92)	0.92 (0.89-0.95)	0.87 (0.84-0.89)
9 (CA, NV, AZ, HI)	1.03 (0.99-1.07)	0.98 (0.95-1.02)	0.87 (0.85-0.89)	0.81 (0.79-0.82)	0.95 (0.93-0.97)
10 (AK, ID, OR, WA)	1.01 (0.94-1.07)	0.79 (0.75-0.84)	0.97 (0.94-1.00)	0.84 (0.81-0.86)	0.98 (0.95-1.01)

<sup>a</sup>Each study period consisted of a unique population and was analyzed separately. Values shown are the adjusted incidence rate ratios for temperature-taking and their associated 95% CIs. Reference groups are listed next to the name of the predictor.

<sup>b</sup>Percentage of population living below the 100% federal poverty level at the census tract level from the 2015-2019 American Community Survey.

<sup>c</sup>Categorized as urban if the census tract was part of an urbanized area of 50,000 or more people based on the 2010 US Census.

<sup>d</sup>Received the thermometer through Kinsa's school distribution and engagement program, FLUency.

<sup>e</sup>Based on ages of profiles associated with the device, with child-only households representing devices where a parent has made profiles for their children but not themselves.

<sup>f</sup>Classified using the Centers for Disease Control and Prevention National Center For Chronic Disease Prevention and Health Promotion Regions (corresponding state abbreviations are in parentheses).

### Percent of Readings With Fever Regression

During all study periods examined, users aged 2-5 and 6-11 years had higher odds of having a feverish reading compared to users aged 19-30 (Table 3). Users aged 31-60 years and users aged over 60 years had significantly lower odds of having a feverish reading during all study periods. Users aged over 60 years had the lowest odds of having a feverish reading during wave 1 of COVID-19, with 0.26 (95% CI 0.25-0.28) times the odds compared to users aged 19-30 years.

Men had significantly increased odds of having a reading that was feverish compared to women during all periods, and this

increased with each subsequent period (Table 3). By wave 3 of COVID-19, men had 27% (95% CI 24%-29%) higher odds of a feverish reading compared to women. Urban users had elevated odds of feverish readings compared to rural users in influenza offseason (OR 1.05, 95% CI 1.01-1.10) and influenza season (OR 1.12, 95% CI 1.09-1.16) (Table 3). This relation shifted during wave 1 of COVID-19 when urban users had 0.90 (95% CI 0.86-0.94) times the odds of feverish readings as rural users. There was no association during wave 2, but by wave 3, urban users again had increased odds of feverish readings compared to rural users (OR 1.18, 95% CI 1.12-1.24).

**Table 3.** Characteristics associated with proportion of feverish readings from multivariable mixed-effects logistic regressions, May 2019-February 2021.

Variables	Odds ratio (95% CI) <sup>a</sup>				
	Offseason 2019 (May 1, 2019, to October 31, 2019)	Flu season (November 1, 2019, to February 2, 2020)	Wave 1 of COVID-19 (February 3, 2020, to May 31, 2020)	Wave 2 of COVID-19 (June 1, 2020, to October 31, 2020)	Wave 3 of COVID-19 (November 1, 2020, to February 28, 2021)
<b>Age (ref: 19-30 years)</b>					
0-1 years	0.56 (0.53-0.58)	0.83 (0.79-0.86)	1.05 (1.00-1.10)	2.03 (1.93-2.14)	2.09 (1.98-2.21)
2-5 years	1.69 (1.61-1.76)	1.61 (1.55-1.67)	1.93 (1.86-2.00)	2.57 (2.46-2.69)	2.01 (1.93-2.10)
6-11 years	1.93 (1.85-2.02)	2.08 (2.00-2.15)	1.80 (1.74-1.87)	1.37 (1.30-1.43)	1.09 (1.04-1.14)
12-18 years	1.33 (1.25-1.41)	1.41 (1.35-1.47)	1.26 (1.20-1.31)	0.95 (0.90-1.01)	0.73 (0.69-0.77)
31-60 years	0.85 (0.82-0.89)	0.83 (0.80-0.86)	0.73 (0.71-0.76)	0.80 (0.77-0.84)	0.93 (0.90-0.97)
>60 years	0.73 (0.66-0.81)	0.53 (0.48-0.58)	0.26 (0.25-0.28)	0.48 (0.45-0.52)	0.65 (0.62-0.69)
Gender, men (ref: women)	1.06 (1.04-1.08)	1.06 (1.05-1.08)	1.06 (1.05-1.08)	1.14 (1.12-1.17)	1.27 (1.24-1.29)
Density, urban (ref: rural) <sup>b</sup>	1.05 (1.01-1.10)	1.12 (1.09-1.16)	0.90 (0.86-0.94)	1.00 (0.95-1.06)	1.18 (1.12-1.24)
<b>Poverty (ref: 0 to &lt;10%)<sup>c</sup></b>					
10% to <20%	1.00 (0.96-1.04)	1.02 (0.99-1.05)	1.14 (1.10-1.19)	1.19 (1.13-1.25)	1.15 (1.10-1.21)
20% to <30%	1.00 (0.95-1.06)	1.05 (1.01-1.10)	1.18 (1.11-1.25)	1.29 (1.20-1.40)	1.26 (1.17-1.35)
≥30%	0.99 (0.92-1.08)	1.02 (0.96-1.08)	0.99 (0.91-1.07)	1.30 (1.17-1.43)	1.26 (1.15-1.38)
FLUency user (ref: non-FLUency) <sup>d</sup>	0.19 (0.17-0.21)	0.80 (0.78-0.83)	1.84 (1.74-1.94)	0.72 (0.66-0.78)	0.59 (0.55-0.63)
<b>Household composition (ref: adult-only)<sup>e</sup></b>					
Child-only	2.21 (2.10-2.33)	1.80 (1.73-1.88)	13.66 (13.04-14.31)	5.83 (5.49-6.20)	3.08 (2.91-3.26)
Multigenerational	1.60 (1.52-1.69)	1.51 (1.45-1.58)	4.36 (4.16-4.57)	2.62 (2.46-2.79)	1.98 (1.86-2.09)
<b>Region (ref: 1 [Northeast])<sup>f</sup></b>					
2 (DC, MD, WV, DE, NJ, PA, VA)	1.00 (0.94-1.07)	0.99 (0.94-1.05)	1.17 (1.10-1.26)	1.34 (1.22-1.46)	1.02 (0.94-1.11)
3 (GA, FL, NC, SC)	1.02 (0.95-1.09)	1.06 (1.00-1.12)	1.25 (1.16-1.34)	1.89 (1.72-2.07)	1.56 (1.43-1.71)
4 (KY, TN, AL, MS)	0.93 (0.85-1.02)	0.89 (0.83-0.95)	1.23 (1.12-1.36)	2.06 (1.81-2.34)	1.77 (1.57-1.99)
5 (IL, WI, IN, MI, MN, OH)	0.85 (0.80-0.91)	0.88 (0.83-0.92)	1.20 (1.13-1.28)	1.24 (1.14-1.34)	1.10 (1.02-1.19)
6 (OK, AR, LA, NM, TX)	0.97 (0.91-1.04)	1.00 (0.95-1.05)	1.27 (1.19-1.37)	1.98 (1.81-2.17)	1.73 (1.59-1.89)
7 (NE, IA, KS, MO)	0.81 (0.74-0.88)	0.87 (0.82-0.93)	1.24 (1.13-1.35)	1.31 (1.17-1.48)	1.37 (1.22-1.54)
8 (MT, ND, WY, CO, SD, UT)	0.81 (0.73-0.90)	0.85 (0.78-0.93)	0.81 (0.72-0.90)	1.12 (0.98-1.28)	1.10 (0.97-1.25)
9 (CA, NV, AZ, HI)	0.99 (0.93-1.05)	1.07 (1.02-1.13)	0.99 (0.93-1.06)	1.57 (1.45-1.70)	1.42 (1.31-1.53)
10 (AK, ID, OR, WA)	0.79 (0.72-0.88)	0.91 (0.84-1.00)	0.58 (0.53-0.64)	0.73 (0.64-0.82)	0.86 (0.76-0.98)

<sup>a</sup>Each study period consisted of a unique population and was analyzed separately. Values shown are the odds ratios with their associated 95% CIs. Reference groups are listed next to the name of the predictor.

<sup>b</sup>Percentage of population living below the 100% federal poverty level at the census tract level from the 2015-2019 American Community Survey.

<sup>c</sup>Categorized as urban if the census tract was part of an urbanized area of 50,000 or more people based on the 2010 US Census.

<sup>d</sup>Received the thermometer through Kinsa's school distribution and engagement program, FLUency.

<sup>e</sup>Based on ages of profiles associated with the device, with child-only households representing devices where a parent has made profiles for their children but not themselves.

<sup>f</sup>Classified using the Centers for Disease Control and Prevention National Center For Chronic Disease Prevention and Health Promotion Regions.

## Discussion

### Principal Findings

Using data collected from smart thermometers, we analyzed temperature-taking behaviors through periods prior to and during the COVID-19 pandemic. We found that both the frequency of readings and proportion of feverish readings varied with age group, gender, urban/rural status, and circulating illness. The differences observed between demographic groups reflect a combination of changes in both actual illness risk and perceived risk that can only fully be understood through dual examination of the number of readings and the percent of those readings with a fever.

Users aged over 60 years experienced the largest shift in temperature-taking behaviors over the study period: during the influenza season, they were less likely to take their temperatures than young adults (aged 19-30 years), whereas through all three waves of COVID-19 assessed in the study, they had an elevated frequency of temperature-taking and reduced proportion of feverish readings.

### Comparison With Prior Work

The significant switch in behavior among older adult users could reflect increases in perceived risk during COVID-19. Adults 65 years and older have higher odds of COVID-19-related concerns [2], and outcomes of hospitalization and death have been the most severe among older adults [27]. Additionally, individuals are more likely to take preventative action if they have a higher perceived risk of a negative health outcome [28]. We hypothesize that the observed shift in temperature-taking behaviors among older adults is related to increased monitoring for signs of possible COVID-19 infection, given its potential severe outcomes, even in the absence of other symptoms. The proportion of older adults among Kinsa users also increased over the three waves of COVID-19 (Table 1), likely as older adults bought and used thermometers more during the pandemic due to increases in perceived risk.

Users aged 2-5 years and 6-11 years had both increased rates of temperature-taking frequency and increased odds of those readings being feverish compared to young adults during all study periods. Child temperature-taking likely reflects parental behavior and concern. Typically, we would expect that as temperature-taking frequency increases, the percent of feverish readings would decrease because the denominator becomes larger. However, our findings suggest that during the COVID-19 pandemic, children were more likely than young adults to have a fever of any origin, as children's temperatures were taken more often and each of those readings had a higher odds of being feverish. Similar to testing for COVID-19, only by examining both the frequency of readings and the percent of those readings with a fever can patterns of behavior and disease be separated [29]. Before the COVID-19 pandemic, children were found to experience febrile illness more often in an average year than adults [16], which may contribute to the observed increased odds of feverish readings. Because children were less likely to experience severe disease from COVID-19 [30], it is possible that child temperature-taking was undermeasured due to reduced perceived risk from caretakers. A reduced risk of

severe outcomes in children may also explain why children were tested for COVID-19 less often than adults [18].

In line with previous research on most personal health behaviors [31,32], men took their temperature less frequently than women across all study periods (Table 2). Lower rates of monitoring likely explain why when men did take their temperature, they had higher odds of being feverish (Table 3). Prior research has found that men were less likely to pay attention to global pandemics than women [33] and less likely to be concerned about COVID-19 during the first two waves of the pandemic [2]. Globally, women have been tested for COVID-19 more frequently than men [34]. Similarly, within the United States, men were tested for COVID-19 less often and had a higher test positivity rate than women [35]. The increased odds of fever we observed in men likely reflects both a decreased perceived risk among men and increased temperature-taking behavior among women.

Urban users had higher rates of temperature-taking than rural users during all study periods, except for wave 2 of the pandemic. Additionally, during wave 1 of COVID-19, researchers found that urban residents reported increases in other health behaviors such as mask wearing and social distancing compared to rural residents [36]. It is likely that either urban users overmonitored their temperatures or rural users undermonitored theirs, since the odds of urban users having feverish readings decreased during wave 1 compared to rural users. Rural areas were more heavily impacted by wave 2 of COVID-19 in terms of cases, hospitalizations, and deaths [37]; accordingly, urban users may have had a decreased perceived risk relative to rural users, leading to their decrease in temperature-taking frequency during this time.

Our study has many strengths that contribute to the literature on surveillance and health behavior. We have a large sample size across multiple illness seasons. Because our behavioral data are not reliant on self-report, we gained an accurate, real-time measure of temperature-taking behavior that is not subject to recall bias. Furthermore, fever data from smart thermometers have been correlated with both COVID-19 cases and influenza-like illness levels [16-19]. Unlike cross-sectional studies that were initiated after the start of the COVID-19 pandemic, we examined data across a previous influenza season, offseason, and multiple waves of COVID-19.

### Limitations

There are also limitations to our study design worth noting. The series of cross-sectional studies do not represent the same population over time and therefore could reflect changes in the user populations rather than changes in user behavior. Future studies should follow a cohort across multiple illness seasons. Because we condition on owning and using a smart thermometer, our study population may not be representative of the US population that does not own a smartphone. However, we capture a wide range of socioeconomic status (Table 1). We also did not have individual-level data on poverty, race, education, or occupation, which could confound some of the observed associations. Additional factors, besides demographics and illness risk, may have impacted the results if a user took a temperature each time they had a possible fever. Users also

could have misassigned fevers to family members if they forgot to switch profiles before taking a temperature.

### Conclusions

Unlike survey data, temperature-taking provides real-time insights into individual behaviors and concerns about circulating infectious disease. Thermometer usage rises with disease circulation, as the highest frequencies were observed during influenza season and wave 1 of COVID-19. Demographic

groups react differently to changes in disease levels, with rural residents and young men taking their temperature less often. These behavioral shifts likely reflect perceived risk more than actual risk. Future studies should investigate how upstream factors such as media coverage impact perceived risk and temperature-taking behavior. Public health surveillance should consider how these behaviors affect testing and health monitoring in interpreting disease levels in different demographic groups.

### Acknowledgments

We thank the members of the Kinsa Data Team for their invaluable feedback during all stages of the analysis.

### Conflicts of Interest

JS and MP are employees of Kinsa, Inc. DB and JZ are employees of Kinsa, Inc, and are shareholders in this company. IS is an employee of Kinsa, Inc; shareholder; and board member of this company.

### Multimedia Appendix 1

Sensitivity analysis examining demographic characteristics associated with temperature-taking frequency measured by the number of days with a reading. Results of the mixed-effects negative binomial regressions for each period. Values shown are the adjusted incidence rate ratios and their associated 95% CIs. Reference groups are listed next to the name of the predictor.

[\[DOCX File, 22 KB - formative\\_v6i9e37509\\_app1.docx\]](#)

### References

- Short S, Mollborn S. Social determinants and health behaviors: conceptual frames and empirical advances. *Curr Opin Psychol* 2015 Oct;5:78-84 [FREE Full text] [doi: [10.1016/j.copsyc.2015.05.002](https://doi.org/10.1016/j.copsyc.2015.05.002)] [Medline: [26213711](https://pubmed.ncbi.nlm.nih.gov/26213711/)]
- Naeim A, Baxter-King R, Wenger N, Stanton A, Sepucha K, Vavreck L. Effects of age, gender, health status, and political party on COVID-19-related concerns and prevention behaviors: results of a large, longitudinal cross-sectional survey. *JMIR Public Health Surveill* 2021 Apr 28;7(4):e24277 [FREE Full text] [doi: [10.2196/24277](https://doi.org/10.2196/24277)] [Medline: [33908887](https://pubmed.ncbi.nlm.nih.gov/33908887/)]
- Ovrum A, Gustavsen G, Rickertsen K. Age and socioeconomic inequalities in health: examining the role of lifestyle choices. *Adv Life Course Res* 2014 Mar;19:1-13. [doi: [10.1016/j.alcr.2013.10.002](https://doi.org/10.1016/j.alcr.2013.10.002)] [Medline: [24796874](https://pubmed.ncbi.nlm.nih.gov/24796874/)]
- Xu J, Peng Z. People at risk of influenza pandemics: the evolution of perception and behavior. *PLoS One* 2015;10(12):e0144868 [FREE Full text] [doi: [10.1371/journal.pone.0144868](https://doi.org/10.1371/journal.pone.0144868)] [Medline: [26658371](https://pubmed.ncbi.nlm.nih.gov/26658371/)]
- Biehl S, Schmidmeier M, Wechsler T, Kroczyk L, Mühlberger A. Widespread recommendations can change our habits of hand-washing and physical distance during the COVID-19 pandemic. *Clin Psychol Eur* 2021 Mar 10;3(1):e3061. [doi: [10.32872/cpe.3061](https://doi.org/10.32872/cpe.3061)]
- Zhang C, Zhang R, Schwarzer R, Hagger MS. A meta-analysis of the health action process approach. *Health Psychol* 2019 Jul;38(7):623-637. [doi: [10.1037/hea0000728](https://doi.org/10.1037/hea0000728)] [Medline: [30973747](https://pubmed.ncbi.nlm.nih.gov/30973747/)]
- Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors. *Eur Psychol* 2008 Jan;13(2):141-151. [doi: [10.1027/1016-9040.13.2.141](https://doi.org/10.1027/1016-9040.13.2.141)]
- Reuter T, Renner B. Who takes precautionary action in the face of the new H1N1 influenza? Prediction of who collects a free hand sanitizer using a health behavior model. *PLoS One* 2011;6(7):e22130 [FREE Full text] [doi: [10.1371/journal.pone.0022130](https://doi.org/10.1371/journal.pone.0022130)] [Medline: [21789224](https://pubmed.ncbi.nlm.nih.gov/21789224/)]
- Clements J. Knowledge and behaviors toward COVID-19 among US residents during the early days of the pandemic: cross-sectional online questionnaire. *JMIR Public Health Surveill* 2020 May 08;6(2):e19161 [FREE Full text] [doi: [10.2196/19161](https://doi.org/10.2196/19161)] [Medline: [32369759](https://pubmed.ncbi.nlm.nih.gov/32369759/)]
- Hamilton K, Smith S, Keech J, Moyers S, Hagger M. Application of the health action process approach to social distancing behavior during COVID-19. *Appl Psychol Health Well Being* 2020 Dec;12(4):1244-1269 [FREE Full text] [doi: [10.1111/aphw.12231](https://doi.org/10.1111/aphw.12231)] [Medline: [33006814](https://pubmed.ncbi.nlm.nih.gov/33006814/)]
- Teasdale E, Santer M, Geraghty AWA, Little P, Yardley L. Public perceptions of non-pharmaceutical interventions for reducing transmission of respiratory infection: systematic review and synthesis of qualitative studies. *BMC Public Health* 2014 Jun 11;14(1):589 [FREE Full text] [doi: [10.1186/1471-2458-14-589](https://doi.org/10.1186/1471-2458-14-589)] [Medline: [24920395](https://pubmed.ncbi.nlm.nih.gov/24920395/)]
- Carroll N, Sadowski A, Laila A, Hruska V, Nixon M, Ma D, on behalf of the Guelph Family Health Study. The impact of COVID-19 on health behavior, stress, financial and food security among middle to high income Canadian families with young children. *Nutrients* 2020 Aug 07;12(8):2352 [FREE Full text] [doi: [10.3390/nu12082352](https://doi.org/10.3390/nu12082352)] [Medline: [32784530](https://pubmed.ncbi.nlm.nih.gov/32784530/)]



13. Cindrich S, Lansing J, Brower C, McDowell C, Herring M, Meyer J. Associations between change in outside time pre- and post-COVID-19 public health restrictions and mental health: brief research report. *Front Public Health* 2021;9:619129. [doi: [10.3389/fpubh.2021.619129](https://doi.org/10.3389/fpubh.2021.619129)] [Medline: [33585393](https://pubmed.ncbi.nlm.nih.gov/33585393/)]
14. Knell G, Robertson M, Dooley E, Burford K, Mendez K. Health behavior changes during COVID-19 pandemic and subsequent "Stay-at-Home" orders. *Int J Environ Res Public Health* 2020 Aug 28;17(17):1 [FREE Full text] [doi: [10.3390/ijerph17176268](https://doi.org/10.3390/ijerph17176268)] [Medline: [32872179](https://pubmed.ncbi.nlm.nih.gov/32872179/)]
15. Privacy policy. Kinsa. URL: <https://kinsahealth.com/privacy-policy> [accessed 2021-09-10]
16. Miller A, Singh I, Koehler E, Polgreen P. A smartphone-driven thermometer application for real-time population- and individual-level influenza surveillance. *Clin Infect Dis* 2018 Jul 18;67(3):388-397. [doi: [10.1093/cid/ciy073](https://doi.org/10.1093/cid/ciy073)] [Medline: [29432526](https://pubmed.ncbi.nlm.nih.gov/29432526/)]
17. Ackley SF, Pilewski S, Petrovic VS, Worden L, Murray E, Porco TC. Assessing the utility of a smart thermometer and mobile application as a surveillance tool for influenza and influenza-like illness. *Health Informatics J* 2020 Sep 23;26(3):2148-2158 [FREE Full text] [doi: [10.1177/1460458219897152](https://doi.org/10.1177/1460458219897152)] [Medline: [31969046](https://pubmed.ncbi.nlm.nih.gov/31969046/)]
18. Miller AC, Peterson RA, Singh I, Pilewski S, Polgreen PM. Improving state-level influenza surveillance by incorporating real-time smartphone-connected thermometer readings across different geographic domains. *Open Forum Infect Dis* 2019;6(11):ofz455. [doi: [10.1093/ofid/ofz455](https://doi.org/10.1093/ofid/ofz455)]
19. Kogan N, Clemente L, Liautaud P, Kaashoek J, Link N, Nguyen A, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. *Sci Adv* 2021 Mar;7(10):1 [FREE Full text] [doi: [10.1126/sciadv.abd6989](https://doi.org/10.1126/sciadv.abd6989)] [Medline: [33674304](https://pubmed.ncbi.nlm.nih.gov/33674304/)]
20. FLUency School Health Program. Kinsa. 2021. URL: <https://kinsahealth.com/kinsa-for-schools> [accessed 2021-09-10]
21. Improving Basic Programs Operated by Local Educational Agencies (ESEA Title I, Part A). US Department of Education Office of Elementary and Secondary Education. 2018. URL: <https://tinyurl.com/5b3bukcy> [accessed 2022-07-10]
22. Estimated flu-related illnesses, medical visits, hospitalizations, and deaths in the United States — 2019–2020 flu season. Centers for Disease Control and Prevention. 2021 Oct 01. URL: <https://www.cdc.gov/flu/about/burden/2019-2020.html> [accessed 2022-08-25]
23. American Community Survey 5-year data (2009-2019). United States Census Bureau. URL: <https://www.census.gov/data/developers/data-sets/acs-5year.html> [accessed 2021-09-10]
24. National Center For Chronic Disease Prevention and Health Promotion Regions. Centers for Disease Control and Prevention. 2013. URL: <https://www.cdc.gov/coordinatedchronic/docs/nccdphp-regions-map.pdf> [accessed 2021-09-10]
25. 2010 Census urban and rural classification and urban area criteria. United States Census Bureau. 2010. URL: <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html> [accessed 2021-09-10]
26. Brooks M, Kristensen K, Benthem K, Magnusson A, Berg C, Nielsen A, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 2017;9(2):378. [doi: [10.32614/rj-2017-066](https://doi.org/10.32614/rj-2017-066)]
27. Cannistraci C, Valsecchi M, Capua I. Age-sex population adjusted analysis of disease severity in epidemics as a tool to devise public health policies for COVID-19. *Sci Rep* 2021 Jun 03;11(1):11787. [doi: [10.1038/s41598-021-89615-4](https://doi.org/10.1038/s41598-021-89615-4)] [Medline: [34083555](https://pubmed.ncbi.nlm.nih.gov/34083555/)]
28. Chapman G, Coups E. Emotions and preventive health behavior: worry, regret, and influenza vaccination. *Health Psychol* 2006 Jan;25(1):82-90 [FREE Full text] [doi: [10.1037/0278-6133.25.1.82](https://doi.org/10.1037/0278-6133.25.1.82)] [Medline: [16448301](https://pubmed.ncbi.nlm.nih.gov/16448301/)]
29. Jamshidi B, Bekrizadeh H, Zargarani S, Rezaei M. Analysis of the number of tests, the positivity rate, and their dependency structure during COVID-19 pandemic. *MedRxiv*. 2021 Apr 23. URL: <https://medrxiv.org/cgi/content/short/2021.04.20.21255796> [accessed 2021-09-10]
30. Perez-Saez J, Lauer S, Kaiser L, Regard S, Delaporte E, Guessous I, Serocov-POP Study Group. Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. *Lancet Infect Dis* 2021 Apr;21(4):e69-e70 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30584-3](https://doi.org/10.1016/S1473-3099(20)30584-3)] [Medline: [32679085](https://pubmed.ncbi.nlm.nih.gov/32679085/)]
31. Dawson K, Schneider M, Fletcher P, Bryden P. Examining gender differences in the health behaviors of Canadian university students. *J R Soc Promot Health* 2007 Jan;127(1):38-44. [doi: [10.1177/1466424007073205](https://doi.org/10.1177/1466424007073205)] [Medline: [17319316](https://pubmed.ncbi.nlm.nih.gov/17319316/)]
32. Rehman R, Zafar A, Mohib A, Baig M. A gender-based comparison in health behaviors and state of happiness among university students. *Cureus* 2018 Mar 19;10(3):e2342 [FREE Full text] [doi: [10.7759/cureus.2342](https://doi.org/10.7759/cureus.2342)] [Medline: [29796354](https://pubmed.ncbi.nlm.nih.gov/29796354/)]
33. Ek S. Gender differences in health information behaviour: a Finnish population-based survey. *Health Promot Int* 2015 Sep;30(3):736-745. [doi: [10.1093/heapro/dat063](https://doi.org/10.1093/heapro/dat063)] [Medline: [23985248](https://pubmed.ncbi.nlm.nih.gov/23985248/)]
34. The COVID-19 sex disaggregated data tracker. The Sex, Gender and COVID-19 Project. 2021. URL: <https://globalhealth5050.org/the-sex-gender-and-covid-19-project/the-data-tracker/?explore=country> [accessed 2021-09-10]
35. Scully E, Schumock G, Fu M, Massaccesi G, Muschelli J, Betz J, et al. Sex and gender differences in testing, hospital admission, clinical presentation, and drivers of severe outcomes from COVID-19. *Open Forum Infect Dis* 2021 Sep;8(9):ofab448 [FREE Full text] [doi: [10.1093/ofid/ofab448](https://doi.org/10.1093/ofid/ofab448)] [Medline: [34584899](https://pubmed.ncbi.nlm.nih.gov/34584899/)]
36. Luo Y, Yao L, Hu L, Zhou L, Yuan F, Zhong X. Urban and rural disparities of personal health behaviors and the influencing factors during the COVID-19 outbreak in China: based on an extended IMB model. *Disaster Med Public Health Prep* 2020 Nov 19:1-5 [FREE Full text] [doi: [10.1017/dmp.2020.457](https://doi.org/10.1017/dmp.2020.457)] [Medline: [33208220](https://pubmed.ncbi.nlm.nih.gov/33208220/)]

37. Cuadros D, Branscum A, Mukandavire Z, Miller F, MacKinnon N. Dynamics of the COVID-19 epidemic in urban and rural areas in the United States. *Ann Epidemiol* 2021 Jul;59:16-20 [FREE Full text] [doi: [10.1016/j.annepidem.2021.04.007](https://doi.org/10.1016/j.annepidem.2021.04.007)] [Medline: [33894385](https://pubmed.ncbi.nlm.nih.gov/33894385/)]

## Abbreviations

**HAPA:** Health Action Process Approach

**IRR:** incidence rate ratio

**OR:** odds ratio

*Edited by A Mavragani; submitted 23.02.22; peer-reviewed by A Carnahan, J Reis; comments to author 16.06.22; revised version received 27.07.22; accepted 19.08.22; published 08.09.22.*

*Please cite as:*

*Seifarth J, Pinaire M, Zicker J, Singh I, Bloch D*

*Circulating Illness and Changes in Thermometer Use Behavior: Series of Cross-sectional Analyses*

*JMIR Form Res* 2022;6(9):e37509

URL: <https://formative.jmir.org/2022/9/e37509>

doi: [10.2196/37509](https://doi.org/10.2196/37509)

PMID: [35998174](https://pubmed.ncbi.nlm.nih.gov/35998174/)

©Jack Seifarth, Megan Pinaire, John Zicker, Inder Singh, Danielle Bloch. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 08.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Optimizing the Acceptability, Adherence, and Inclusiveness of the COVID Radar Surveillance App: Qualitative Study Using Focus Groups, Thematic Content Analysis, and Usability Testing

Bas Splinter<sup>1</sup>, MD; Nicholas H Saadah<sup>1,2</sup>, PhD; Niels H Chavannes<sup>1,2</sup>, MD, PhD; Jessica C Kiefte-de Jong<sup>1,2</sup>, PhD; Jiska J Aardoom<sup>1,2</sup>, PhD

<sup>1</sup>Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

<sup>2</sup>National eHealth Living Lab, Leiden, Netherlands

**Corresponding Author:**

Jiska J Aardoom, PhD

Department of Public Health and Primary Care

Leiden University Medical Center

Hippocratespad 21

Leiden, 2300 RC

Netherlands

Phone: 31 071 526 843

Email: [J.J.Aardoom@lumc.nl](mailto:J.J.Aardoom@lumc.nl)

## Abstract

**Background:** The COVID Radar app was developed as a population-based surveillance instrument to identify at-risk populations and regions in response to the COVID-19 pandemic. The app boasts of >8.5 million completed questionnaires, with >280,000 unique users. Although the COVID Radar app is a valid tool for population-level surveillance, high user engagement is critical to the success of the COVID Radar app in maintaining validity.

**Objective:** This study aimed to identify optimization targets of the COVID Radar app to improve its acceptability, adherence, and inclusiveness.

**Methods:** The main component of the COVID Radar app is a self-report questionnaire that assesses COVID-19 symptoms and social distancing behaviors. A total of 3 qualitative substudies were conducted. First, 3 semistructured focus group interviews with end users (N=14) of the app were conducted to gather information on user experiences. The output was transcribed and thematically coded using the framework method. Second, a similar qualitative thematic analysis was conducted on 1080 end-user emails. Third, usability testing was conducted in one-on-one sessions with 4 individuals with low literacy levels.

**Results:** All 3 substudies identified optimization targets in terms of design and content. The results of substudy 1 showed that the participants generally evaluated the app positively. They reported the app to be user-friendly and were satisfied with its design and functionalities. Participants' main motivation to use the app was to contribute to science. Participants suggested adding motivational tools to stimulate user engagement. A larger national publicity campaign for the app was considered potentially helpful for increasing the user population. In-app updates informing users about the project and its outputs motivated users to continue using the app. Feedback on the self-report questionnaire, stemming from substudies 1 and 2, mostly concerned the content and phrasing of the questions. Furthermore, the section of the app allowing users to compare their symptoms and behaviors to those of their peers was found to be suboptimal because of difficulties in interpreting the figures presented in the app. Finally, the output of substudy 3 resulted in recommendations primarily related to simplification of the text to render it more accessible and comprehensible for individuals with low literacy levels.

**Conclusions:** The convenience of app use, enabling personal adjustments of the app experience, and considering motivational factors for continued app use (ie, altruism and collectivism) were found to be crucial to procuring and maintaining a population of active users of the COVID Radar app. Further, there seems to be a need to increase the accessibility of public health tools for individuals with low literacy levels. These results can be used to improve the this and future public health apps and improve the representativeness of their user populations and user engagement, ultimately increasing the validity of the tools.

(JMIR Form Res 2022;6(9):e36003) doi:[10.2196/36003](https://doi.org/10.2196/36003)

**KEYWORDS**

COVID-19; corona; eHealth; self-report; mobile app; track-and-trace strategies; population surveillance; citizen science; usability; mobile phone

## Introduction

### Background

Since December 2019, the world has been battling the SARS-CoV-2 (COVID-19). In the Netherlands specifically, the first case was identified in February 2020 and has resulted, to date, in >2.6 million confirmed cases and 19,000 deaths [1]. Track-and-trace strategies and quarantine measures to limit social contact and social distancing have been widely used to prevent the transmission of the virus [2]. However, a delay between the appearance of symptoms and confirmed test results increase lead times during an outbreak. Continuous population-based monitoring of COVID-19-related symptoms and social distancing behaviors may aid in the estimation of risk of COVID-19 cases at a regional level, allowing local governments to intervene at an earlier stage [3].

Since the start of the pandemic, mobile apps have been implemented for a variety of goals such as risk assessment and decision-making as well as self-management and self-monitoring of symptoms [4]. Symptom-monitoring apps based on self-reporting show promising results in terms of predicting local COVID-19 spread using symptom-based tracking in the United Kingdom and United States [5,6]. Menni et al [5] found that a prediction model based on an app-based symptom tracker could predict a positive COVID-19 test with a sensitivity of 65% and a specificity of 78% [5]. Other studies have shown that population-wide self-reported data collected by means of a mobile app could identify 75% of the regions with the highest COVID-19 incidence according to governmental test data [6]. Finally, a recent study showed that longitudinal self-reported data on health, behavior, and demographics, as collected with web and mobile apps, could be adequately used in building prediction models to identify potential positive COVID-19 cases [7]. The above literature highlights the utility of self-reported data in population-based COVID-19 tracking and subsequent prediction of COVID-19 hot spots, which can ultimately aid in the response to the COVID-19 pandemic.

In the Netherlands, a smartphone-based surveillance app, the COVID Radar app, was developed. The goal of the app was to predict topographical future COVID hot spots with the help of self-reporting of COVID symptoms, enabling the possibility for local or national governments to intervene, for instance, by informing the public or implementing local restrictions. The app enables frequent and anonymous voluntary self-reporting of COVID-19-related symptoms and behaviors (users are asked only for their postcode, age range, sex, and profession). The app was developed during the first COVID-19 wave in the Netherlands in the spring of 2020. Among other functionalities, the app contains a short monitoring questionnaire ( $\geq 20$  questions) asking users to self-report their COVID-19-related symptoms, social distancing behaviors, COVID-19 status, and vaccination status. The app additionally provides users feedback on their reported social distancing behavior (eg, number of

people encountered within 1.5 meters and hours spent outside the house) and symptoms by enabling them to compare these to the mean values nationwide as well as within their specific geographic region. Since the launch of the app in April 2020, it boasts of >8.5 million completed questionnaires, with >280,000 unique users, of whom >13,000 completed the questionnaire at least twice a week on average. The COVID Radar app can be considered a citizen science project, as it involves the public in the scientific processes, in this case the collection of large-volume longitudinal symptom and social distancing data across the Netherlands [8].

Previous research has shown that the COVID Radar app is a useful and valid tool for population-level surveillance and potentially for the prediction of local COVID-19 hot spots [3]. More specifically, self-reported positive test results reported via the app closely matched government-reported case counts. In addition, there were clear associations between self-reported COVID-19 symptoms and positive test results. With respect to behavioral measures, a clear association was found between self-reported positive test results and above-average risk behaviors in terms of social distancing (eg, having had more visitors in one's home or having had more people within 1.5 meters) in the days leading up to a test. Importantly, the identified associations among symptoms, social distancing behaviors, and test results were most pronounced in areas with high user engagement. Hence, high user engagement throughout different regions of a country seems critical to the success of any predictive model using self-reported symptoms and behavior. Stimulating and increasing user engagement by citizens in population-based surveillance apps such as the COVID Radar remains challenging.

The Netherlands Organization of Health Research and Development funded the COVID Radar project in which a multidisciplinary team aimed to develop a mobility- and behavior-based early warning system after the first wave of COVID-19 in the Netherlands. The methods in this project combine unique real-time spatial summaries per 4-digit postal code area of symptoms and high-risk behavior from the population surveillance data of the COVID Radar app at the Leiden University Medical Center (LUMC), with aggregated historic mobility information provided by mobile telecommunications data. One of the objectives of the project is to optimize the COVID Radar app and enrich the population surveillance syndrome and behavior data. The goal of this optimization of the COVID Radar app and survey is to ultimately improve the acceptability, adherence, and inclusiveness of the app and to learn lessons for improvement of future population-level health-related apps. This can subsequently support further national upscaling and use of the COVID Radar, as well as improve the representativeness of its user population. Furthermore, such targets can be subsequently used as inputs when designing national surveillance self-report data collection apps.

## Objectives

The main aim of this study was to identify the optimization targets of the COVID Radar app to ultimately improve the acceptability, adherence, and inclusiveness of the app. We aimed to identify optimization targets by conducting 3 qualitative substudies: (1) gathering and analyzing in-depth information on user experiences by means of semistructured focus group interviews with end users of the COVID Radar app, (2) analyzing all received end-user emails that were sent to the COVID Radar project team since the launch of the app, and (3) review by language experts and usability testing of the COVID Radar app in individuals with low literacy levels.

## Methods

### Design

All 3 substudies were qualitative research investigations. The first substudy consisted of 3 semistructured focus group interviews with COVID Radar end users. The second substudy comprised a qualitative analysis of all emails sent by the end users to the COVID Radar project team. The third substudy involved expert reviews and individual usability test sessions with individuals with low literacy levels. More details on the methods of each of these substudies can be found below.

### COVID Radar App

The COVID Radar app was freely available for iOS and Android systems and could be downloaded from both the Apple App Store and Google Play Store between April 2, 2020, and February 28, 2022. Upon its release, a brief publicity campaign was launched. First-time users are asked to register by providing information about (1) their gender (male, female, other, or not specified), (2) their age (10-year intervals from 0 to 80 and a  $\geq 80$  category), (3) their occupation (education, health care, catering industry, or other occupations with high risk of human contact within 1.5 meters), and (4) the 4 digits of their postal code.

The app consists of 4 sections. The first section contains the self-report monitoring questionnaire. A screenshot of the questionnaire is presented in [Figure 1](#). The questionnaire is dynamic, meaning that the number and content of questions can be updated as needed based on new relevant insights related to the surveillance of the coronavirus. For example, questions about vaccination status were added as soon as the Dutch

vaccination program started. The final questionnaire contained a total of 23 questions. The full version of the current questionnaire is presented in [Multimedia Appendix 1](#). A total of 11 questions assess COVID-19 symptoms (eg, coughing, fever, and sore throat), and 8 questions assess social distancing behaviors, including among others, the number of human contacts within 1.5 meters, whether one has been in contact with a patient with COVID-19 infection in the last 14 days, how many visitors one had received on a particular day, whether the participant was ever tested positive (yes or no), whether the participant was tested in the last 2 weeks (no, have not been tested; have been tested—result was that I did not have the coronavirus; have been tested, result was that I did have the coronavirus), and a question about COVID vaccine status. Users receive a push message every other day to remind them to fill in the questionnaire, regardless of whether the user has already done so.

The second section of the COVID Radar app is an update section. This section is used to inform users on changes in the self-report questionnaire or to inform them about recent results of analyses of COVID Radar data.

The third section contains a frequently asked questions (FAQs) section, where users can find information on numerous topics related to information about the app (eg, the goal of the app), data protection (eg, which data are gathered exactly, and why will the data be saved for a period of 5 years?), contact information, and instructions on how to complete the questionnaire (eg, how to report symptoms when having asthma?).

The fourth and final section is the Radar section, where feedback on user-reported symptoms and social distancing behavior can be observed. More specifically, app users can compare their self-reported social distancing behaviors (eg, the number of people spoken to within 1.5 meters) and symptoms to those of other users in their local region as well as the country's app users as a whole. The feedback regarding social distancing behaviors is provided using sliders, colored from green to brown, with green representing *safe* behavior and red representing relatively *risky* behavior. A screenshot is provided in [Figure 2](#). [Figure 3](#) shows the mean prevalence of symptoms per geographic region, categorized as brown (many people with symptoms), green (few people with symptoms), and gray (insufficient participants to adequately measure the prevalence).

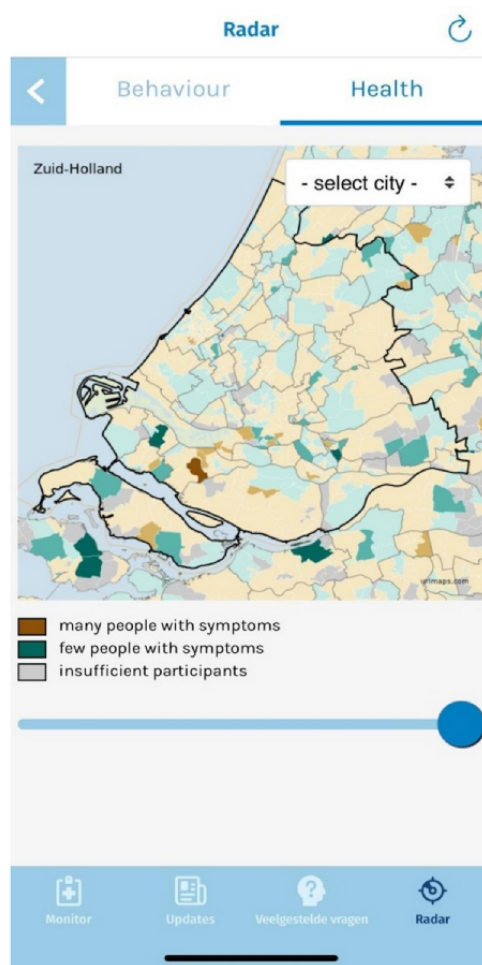
**Figure 1.** Screenshot of the monitoring section, presenting part of the self-report monitoring questionnaire.



**Figure 2.** A screenshot of the Radar section, presenting a map with the mean symptoms per geographic region, as well as sliders providing feedback on the prevalence of symptoms compared with other users.



**Figure 3.** A screenshot of the Radar section, presenting a map with mean social distancing behaviors per geographic region, as well as sliders providing feedback on social distancing behavior in comparison with other users.



## Substudy 1: Focus Group Interviews With End Users

### Procedure

COVID Radar app users were notified of the upcoming focus groups and invited to participate via an update in the update section of the app. The invitation contained a brief summary of the study aims and corresponding procedures. For participating in one of the focus groups, an individual had to be able to read, speak, and understand Dutch. Participants could indicate their interest in participating by sending an email to the research team. A group of 50 potential participants was randomly chosen from among the responses, who were subsequently asked to provide their availability for the upcoming focus groups. A total of 3 focus groups were scheduled to accommodate varying schedules. Because of possible technical issues, the focus groups were conducted with smaller groups than the offline focus group, as recommended by previous research [9,10]. A total of 5 available participants were chosen at random for each focus group.

Two weeks before the focus group would take place, participants received an information letter by email containing all study details and corresponding focus group procedures. The email also contained an informed consent form. Individuals were instructed to complete the form digitally or manually (ie, print, sign, and scan) and to return the completed form to the

researchers by email. Participants received a €20 (US \$24) gift card for their participation in the interviews.

### Focus Group Interviews

A semistructured interview protocol was developed to serve as a guideline for the discussion of relevant topics during the focus group interviews. The interview protocol comprised open questions aimed at gathering participants' experiences and opinions with respect to 7 topics: (1) general opinion of the app, (2) suggestions for improvement of the app, (3) feedback on the self-report monitoring questionnaire, (4) feedback on the update section, (5) feedback on the FAQs section, (6) feedback on the Radar section, and (7) feedback on the push notifications.

A total of 3 focus groups were conducted in May 2021. All focus group interviews were web-based, conducted via Zoom (Zoom Video Communications) with the audio being recorded. The interviews were moderated by BS. NHS and JJA were present as backup moderators and for support in case of technical problems. The duration of each focus group was approximately 60 minutes.

Audio recordings were transcribed literally. The names of participants were replaced with participant numbers to preserve their anonymity. Subsequently, the transcripts were thematically coded using the framework method [11]. The framework is a qualitative content analysis method. It was used to identify

commonalities and differences in qualitative data before exploring the coherence of the data, focusing on finding explanatory conclusions per theme. The framework method works in different stages: (1) transcription, (2) familiarization with transcript, (3) coding, (4) developing an analytical framework, (5) applying the analytical framework, (6) charting data into the framework matrix, and (7) interpreting the data. During the process, it was possible to go back and forth between the different stages. This dynamic approach enabled the addition of new codes during the analysis, thereby creating an opportunity to react to unexpected findings and a broad spectrum of input on themes. This method is often used to analyze semistructured interview data. The identified themes were (1) general opinion of the app, (2) motivation for use, (3) publicity of the app and enlarging the user population, (4) self-report monitoring questionnaire, (5) in-app updates, (6) FAQs, (7) Radar section, and (8) push notifications. Coding was performed by BS and checked for reliability by JS by checking random samples of the coded transcripts. Coding was performed using Atlas.ti software (version 7.5.18). Data saturation was reached after conducting 3 focus group interviews.

### Substudy 2: Email Box Analysis

The COVID Radar team offered the possibility of app users sending feedback to a general email address. This email address was published in the FAQs section under the contact information. The goal was to enable the research team to subsequently optimize the app or handle technical errors as soon as possible. A total of 1080 emails received by the COVID Radar team between May 2020 and June 2021 were exported into a text document and subsequently uploaded to Atlas.ti software (version 7.5.18). The codebook designed for substudy 1 was used for thematic coding of the email feedback.

### Substudy 3: Review by Language Experts and Usability Testing in Individuals With Low Literacy Levels

In order to evaluate the COVID Radar app as a tool for individuals with low literacy levels specifically, 2 types of activities were carried out by Pharos: the Dutch center of expertise on health disparities. The first activity was an expert review of the self-report monitoring questionnaire by 2 experts from Pharos. All the text was reviewed, considering that the language needed to meet language level A2 or B1 according to the Common European Framework of Reference for Languages

[12]. They also made suggestions for adapting the text based on their checklist for accessible information. The checklist contains relevant information related to the accessibility of information in terms of, for example, layout (eg, ensuring there is sufficient color contrast between the text and the background), dosage of information (eg, stating the most important message at the beginning and repeating it), and phrasing and readability of text (eg, avoiding or explaining difficult technical terms or medical jargon). On the basis of the results of the expert review, the researchers updated the phrasing of the questions in the self-report monitoring questionnaire.

The second activity was the usability testing of 4 individuals with low literacy levels. In a one-on-one web-based setting, 4 participants joined a 1-hour usability testing meeting with an expert from Pharos. During this usability testing, the participants were asked to perform assignments to check whether the navigation was simple and logical. They were also asked to read the text aloud so that it became clear which words or sentences were difficult to read. In addition, the participants were asked to explain the meaning of the text in their own words. Finally, the participants were asked for feedback about the entire usability testing process by asking them for alternatives or possible solutions to encountered difficulties. The results of the usability testing were summarized and presented by Pharos to the involved researchers.

### Ethics Approval

Substudies 1 and 2 were declared as not falling under the scope of the Dutch Medical Research Involving Human Subjects Act by the medical ethical committee of the LUMC and were granted a certificate of no objection accordingly. Substudy 3 was certified and performed by a third-party organization (Pharos). All volunteers in these studies signed informed consent forms before their participation.

## Results

### Overview

The codebook developed and used for substudies 1 and 2 is presented in [Table 1](#). The 8 themes were (1) user friendliness, (2) motivation for use, (3) publicity and enlarging the user population, (4) monitoring questionnaire, (5) in-app updates, (6) FAQs, (7) Radar section, and (8) push notifications.



**Table 1.** The codebook used for coding the transcripts of the focus groups (substudy 1) and the email feedback (substudy 2).

Theme and codes	Description	Examples
<b>Theme 1: general opinion of the app</b>		
User friendliness	Comments related to the user friendliness of the app; for example, the ease of using the app and navigating through the app	<ul style="list-style-type: none"> <li>“I find it very down-to-earth, simple, easy to use”</li> <li>“It is a pleasant app to use, but it is indeed very text-heavy. No symbols are used, so you will mainly reach the people with high literacy levels.”</li> </ul>
Design	Comments related to the design of the app, such as colors and layout	<ul style="list-style-type: none"> <li>“I’m just wondering here, but not all colors may be easy to distinguish on mobile phones, because of the often used colors dark blue, light blue, and gray. On many mobile phones it is not easy to distinguish them.”</li> </ul>
Theme 2: motivation for use	Comments related to participants’ motivation and reasons for using the app	<ul style="list-style-type: none"> <li>“I started using the COVID radar to help you gain insight...”</li> <li>“It is not only an app that is relevant to the field of research but also relevant to oneself. You can also use it to monitor your own behavior and the behavior of your environment. So it also holds up a mirror to oneself.”</li> </ul>
Theme 3: publicity of the app and enlarging the user population	Comments related to increasing the user population of the app and supporting the national awareness of the app	<ul style="list-style-type: none"> <li>“Almost all of us [referring to focus group participants] have a connection with Leiden or the Leiden University Medical Center. I live in Leiden and the only newspaper in which I read something about the app was a local Leiden newspaper, so more national publicity may persuade other people to start using your app.”</li> </ul>
<b>Theme 4: self-report monitoring questionnaire</b>		
Content	Comments related to the content of the questions, such as language issues, understandability, suggestions for additional questions, and other suggestions for improvements related to the self-report monitoring questionnaire	<ul style="list-style-type: none"> <li>“It is impossible to estimate how many people come within five meters when you are at something like a market or a supermarket.”</li> <li>“I think the question is how many hours were you out of the house yesterday. Of course it is clear in itself. But the explanation says that a holiday destination counts as leaving home. Well, at one point we spent a week in a house a while ago...I find that a bit strange that I was out of the house that whole week, 24 hours a day. I don’t think that makes sense.”</li> </ul>
Length and duration	Comments related to the time it took to fill in the questionnaire or about the number of questions	<ul style="list-style-type: none"> <li>“And what strikes me is that more and more questions are added and I understand that you want to know more about vaccination and everything, but the list is getting very long.”</li> </ul>
Additional information	Comments related to the use of the icons for additional information or elaboration of questions as well as its content	<ul style="list-style-type: none"> <li>“Yeah in the beginning I clicked on the icons and read their content, you know. Because at one point I wondered ‘yes what about that?’: do I need to count being in the garden as time outside of the house? Well, that didn’t appear to be the case.”</li> </ul>
Frequency of completion	Comments related to the frequency of filling in the questionnaire or about when and where participants completed the questionnaire	<ul style="list-style-type: none"> <li>“Usually, I fill it in at night when I’m on the couch. Often then I think: ‘oh yeah that radar.’”</li> <li>“Every day! I’m one hundred percent sure. It’s almost something compulsive, that I feel that I need to complete it, I must not forget.”</li> </ul>
<b>Theme 5: in-app updates</b>		
Frequency	Comments related to about the frequency or timing for the updates	<ul style="list-style-type: none"> <li>“I think those updates are great, just like what participant X just said; it could be as often as weekly.”</li> </ul>
Effect	Comments related to the experienced effects of in-app updates	<ul style="list-style-type: none"> <li>“They mainly motivate me. Then I think, oh well, nice to hear something back. I also think because I have been using the app for more than a year now, I then also see what I am doing it for. So therefore yes I always read the updates.”</li> </ul>

Theme and codes	Description	Examples
Content	Comments related to the content of the updates	<ul style="list-style-type: none"> <li>“Generally, I find the updates informative and it’s good to have some more background information about the whole subject.”</li> </ul>
<b>Theme 6: FAQs<sup>a</sup></b>		
Reason(s) for consulting	Comments related to (the) reason(s) of why a user consulted the FAQs section and suggestions for improvement	<ul style="list-style-type: none"> <li>“Yes, certainly I looked at the FAQ section. Especially in the beginning before I started filling in the questionnaire. Just to see who is behind this app and what they are doing with my data.”</li> </ul>
Content	Comments related to the content of the FAQs section, such as the clarity of answers and the completeness of topics	<ul style="list-style-type: none"> <li>“This section is clear. Yes, and the answers are too.”</li> </ul>
<b>Theme 7: Radar section</b>		
Content	Comments related to the content of the Radar section (eg, maps and sliders) and its effect	<ul style="list-style-type: none"> <li>“I think you have those sliders regarding behaviour; how many hours have you been outside. It took me a really long time to realize that there are three sliders and not two...”</li> </ul>
Frequency of visiting	Comments related to how often a user takes a look at the Radar section	<ul style="list-style-type: none"> <li>“Yes, I look at it daily. Or daily, every time I fill in the list, that’s not necessarily daily.”</li> </ul>
<b>Theme 8: push notifications</b>		
Suggested improvements	Comments related to suggested ways to improve the push notifications, for example with regard to the timing and frequency	<ul style="list-style-type: none"> <li>“A personalized pushing message, like for me it is part of my morning routine. When I forget to fill in the questionnaire, I would prefer to hear it [the reminder] in the morning because it happens before nine o’clock or it wouldn’t happen.”</li> <li>“And I think that reminder, whether that is by means of a sound or a pop-up, certainly helps. I have noticed myself... Yes, if I would be able to set a reminder on my own time with a snooze function, then I think the completion score would be much higher.”</li> </ul>
Effect	Comments related to the effect of the (nonintelligent) push messages	<ul style="list-style-type: none"> <li>“Well I have to say, I’m really happy with those reminders you get. Because, I am someone who does not complete the questionnaire routinely during a particular part of the day, unfortunately. So I have noticed that it helps when I receive such a message, like a push message.”</li> <li>“I turned them off, because I thought it was very mean that I completed it [the questionnaire] and still received a reminder.”</li> </ul>

<sup>a</sup>FAQ: frequently asked question.

## Substudy 1: Focus Group Interviews

### Overview

In all, 2 focus groups comprised 5 participants and a focus group comprised 4 participants because of a participant not being able

to join the interview owing to personal circumstances. Of the 14 participants, 12 (86%) were female. Other demographic information was not acquired to preserve the anonymity of the participants. The results of substudy 1 are summarized in [Table 2](#). In the text, a brief explanation of the results is presented, according to the 8 identified themes ([Table 1](#)).

**Table 2.** Results of the focus groups as summarized in strengths and potential improvements.

Theme and subthemes	Strengths	Potential improvements
<b>Theme 1: general opinion of the app</b>		
User friendliness	<ul style="list-style-type: none"> <li>• Simple and intuitive to use</li> <li>• Previous answers being saved as default answers</li> </ul>	<ul style="list-style-type: none"> <li>• For some users, the app was a bit too simplistic, which made the app experience boring</li> </ul>
Design	<ul style="list-style-type: none"> <li>• Basic and simple</li> <li>• Satisfaction with functionalities</li> </ul>	<ul style="list-style-type: none"> <li>• Create a dark mode option</li> <li>• Use of more symbols and less text</li> <li>• More use of effective color contrast</li> </ul>
Theme 2: motivation for use	<ul style="list-style-type: none"> <li>• Enables app users to help advance science and contribute to data collection</li> <li>• Increased awareness of social distancing behavior</li> </ul>	<ul style="list-style-type: none"> <li>• Increased exposure to the research goals and the broader aim of the COVID Radar app</li> <li>• Add motivational tools such as gaming elements</li> </ul>
Theme 3: publicity of the app and enlarging the user population	<ul style="list-style-type: none"> <li>• High engagement of citizens with a direct or indirect connection to the research organization and its city</li> </ul>	<ul style="list-style-type: none"> <li>• More effort in terms of national publicity campaigns to enlarge the national user population</li> <li>• More effort in terms of publicity campaigns specifically in geographic areas with relatively low user engagement</li> </ul>
<b>Theme 4: self-report monitoring questionnaire</b>		
Content	<ul style="list-style-type: none"> <li>• Most questions were deemed as important and logical by the users</li> </ul>	<ul style="list-style-type: none"> <li>• Phrase some questions including their answer categories in a more clear and uniform manner, reducing ambiguity</li> <li>• Adapting or deleting difficult-to-answer questions or questions that cannot always be answered in a reliable manner</li> </ul>
Length and duration	<ul style="list-style-type: none"> <li>• The duration of less than a minute supports regular completion</li> </ul>	<ul style="list-style-type: none"> <li>• Create the opportunity for app users to shorten the questionnaire by introducing more logic (ie, skipping questions, where not applicable)</li> </ul>
Additional information	<ul style="list-style-type: none"> <li>• Overall clear</li> </ul>	<ul style="list-style-type: none"> <li>• Elaborate on the provided explanatory information to increase the understanding of questions and thereby provide more support for selecting the right answer category</li> </ul>
Frequency of completion	<ul style="list-style-type: none"> <li>• Most users completed the questionnaire on a regular basis</li> </ul>	<ul style="list-style-type: none"> <li>• Personalized push notifications to motivate to complete the questionnaire and help to create the habit of filling in the questionnaire regularly</li> <li>• Also refer to potential improvements for theme 2 (motivation for use) as well as theme 5 (in-app updates)</li> </ul>
<b>Theme 5: in-app updates</b>		
Frequency	<ul style="list-style-type: none"> <li>• None</li> </ul>	<ul style="list-style-type: none"> <li>• More frequent updates</li> </ul>
Effect	<ul style="list-style-type: none"> <li>• Encourages and motivates to complete and continue to complete the questionnaire</li> </ul>	<ul style="list-style-type: none"> <li>• None</li> </ul>
Content	<ul style="list-style-type: none"> <li>• Interesting</li> </ul>	<ul style="list-style-type: none"> <li>• Increase the understandability of presented figures</li> </ul>
Theme 6: frequently asked questions—content	<ul style="list-style-type: none"> <li>• Clear and comprehensive</li> </ul>	<ul style="list-style-type: none"> <li>• None</li> </ul>
<b>Theme 7: Radar section</b>		
Content	<ul style="list-style-type: none"> <li>• Section provides insight in symptoms and behaviors compared with peers</li> </ul>	<ul style="list-style-type: none"> <li>• Increase the understandability of the figures by increasing its design and elaborate on and better explain the corresponding functionalities.</li> </ul>
Frequency of visiting	<ul style="list-style-type: none"> <li>• None</li> </ul>	<ul style="list-style-type: none"> <li>• Making the Radar section more visible to users.</li> </ul>

Theme and subthemes	Strengths	Potential improvements
Theme 8: push notifications	<ul style="list-style-type: none"> <li>Valuable; helpful in completing the questionnaire more frequently</li> </ul>	<ul style="list-style-type: none"> <li>In case of continued use of automated notifications, disable notifications when questionnaire is already completed</li> <li>Enable personalization in terms of frequency and timing of the notifications</li> </ul>

### General Opinion of the App

In general, participants evaluated the COVID Radar app positively. They found the app to be user-friendly and intuitive. Most of the participants were pleased with the design and functionality of the app, with its simplicity frequently being listed as one of its strengths:

*I find it very down-to-earth. Simple and easy to use. For me the app is loud and clear, yes, I don't think it could be more easy.*

On the other hand, some participants found the app to be a bit too basic and suggested that the simplicity of the app did not challenge them sufficiently:

*How to reach a different or bigger target audience? I think you will have to fix something in the app itself. I am going to say something ugly, but the app looks superficial.*

Overall, participants appreciated the fact that their most recent answers regarding COVID-19 symptoms were saved as default answers, thereby saving them time the next time they filled in the questionnaire.

### Motivation for Use

Participants' main motivation to use the app can be summarized as their desire to advance science. Participants were motivated to contribute to data collection to help scientists predict future COVID-19 flare-ups:

*I'm very pleased with the app, because it gives me the feeling that I'm contributing to something. Of course, that is a great feeling, especially in this time where you often feel helpless.*

In addition, several participants mentioned that the time spent reflecting on their social distancing behaviors increased their awareness thereof. For example, when reporting the number of people coming within 1.5 meters or the number of visitors received in one's home, if these values were relatively high, participants reported experiencing a sort of discomfort that motivated them to reflect on their behavior:

*And then you think, well, perhaps I could do that differently next time. It didn't lead me to self-report it differently than it actually was, although I did think from time to time "maybe I should report less people." But it has kept me on my toes every day, so in that sense I'm definitely doing it for you [the researchers/science], but I'm also doing it for myself and for my own awareness.*

When discussing the motivation to keep using the app, several participants discussed the potential of integrating motivational tools in the app to increase adherence. Particular suggestions

included repeated exposure to the research goals and the broader aim behind the COVID Radar; for example, by means of in-app updates. Other suggestions related to the potential addition of serious gaming elements, such as enabling participants to collect digital badges based on the number of times they completed the questionnaire:

*But you could also use badges, you know. I have another app in which I earn badges for 5 days, 10 days, and 50 days. But then you have to make some kind of dashboard available in the COVID Radar where one can see if one has completed the questionnaire. I know that I am very motivated by those kinds of streaks.*

### Publicity of the App and Enlarging the User Population

Most participants in the focus groups became aware of the app through a direct or indirect personal connection to the LUMC or the city of Leiden. More efforts in terms of national publicity campaigns using social media were suggested to reach more citizens. Another suggestion was to advertise the app in geographic areas with relatively low user engagement using pamphlets or local media.

### Self-report Monitoring Questionnaire

Most feedback on the self-report monitoring questionnaire concerned the content and phrasing of the questions. Several participants pointed out that some questions left room for interpretation by app users, which may lead to biased answers. More specifically, the question "Yesterday, how many people came within 5 meters of you, outside the house?" received extensive feedback. Most participants considered this question challenging or even too difficult to answer, as they did not feel capable of accurately estimating this number:

*I ride my bike almost daily. How many people do you come across within 5 meters? A street is 5 meters, so every time I ride my bike I am almost sure that I can fill in that 50 people came closer than 5 meters to me.*

The question "Are you completely vaccinated against COVID-19?" also received extensive feedback. Some participants wondered why there was only an interest in whether they were completely vaccinated, instead of allowing users to indicate how many times they had received vaccination.

Regarding the additional information accompanying the questions, the provided information did not always clarify questions or doubts the participants had concerning the respective question. For example, there were numerous questions regarding whether certain locations, such as work offices, shops, and supermarkets, should be categorized as public places.

Concerning the length of the questionnaire, some participants found it too long, although most participants reported that filling in the questionnaire took less than a minute. Relatedly, some participants struggled with their motivation to answer the same questions repeatedly. One suggestion was to enable the possibility of indicating whether the participant experienced changes in COVID symptoms, as these were found to be relatively steady in comparison with behavioral measures (eg, one does not have a sore throat every other day, whereas the number of people coming closer than 5 meters can fluctuate by day). More specifically, participants suggested creating a shortened version of the questionnaire in which the symptom complaints of COVID-19 were combined into one question (eg, “Are you experiencing any COVID symptoms?”) which, if answered positively, resulted in a pull-down menu in which they could indicate which specific symptoms were present.

Most participants reported completing the monitoring questionnaire regularly, with some reporting doing so daily. Most participants reported having made a habit of filling in the questionnaire at a set time each day:

*It needs to become just like brushing your teeth.*

Participants who filled in the questionnaire on a less regular basis seemed to report less consistency in their timing of completion, struggling to create a habit.

### **In-App Updates**

Most participants read the updates as soon as they noticed the in-app notification, whereas others never received or noticed any notifications and checked this section once in a while for updates. Overall, the push messages of the in-app updates were evaluated as being valuable.

Participants were enthusiastic about the content of the updates and described the updates as very interesting. Furthermore, participants elaborated on how these updates motivated them to keep filling in the questionnaire:

*For me, the updates are the most fun part of the app, because it is some kind of reward for completing the questionnaire.*

However, numerous participants mentioned that the figures used in the updates were not always comprehensive or explained in sufficient detail. For example, the color scheme was a source of confusion, with some participants not being able to fully understand the figures. It was also mentioned that the axis labels were sometimes lacking or incomplete. Almost all participants indicated that they preferred more frequent updates.

### **Frequently Asked Questions**

Approximately half of the participants reported having accessed the FAQs section of the app at least once. Reasons for consulting the FAQs section ranged from being interested in the research and corresponding analyses performed with the collected data and wanting to know more about privacy-related issues to curiosity as to how many active app users there were. In general, participants perceived the information in this section to be clear and comprehensible.

### **Radar Section**

For a few participants, the Radar section of the app provided a convenient method for comparing their symptoms and behaviors with those of their peers. They elaborated that they sometimes adjusted their behavior based on peer comparison data:

*I compare myself to other people in my region. I am sensitive to social pressure, so for me sometimes this is a reason to hold myself back on the weekends.*

Nevertheless, the Radar section received the most attention in terms of identified optimization targets. Two participants had never noticed the Radar. The visuals (ie, figures) presented in the Radar section were another frequently mentioned subject of potential improvement. Lack of contrast and nonmatching colors in the legends and figures were described as hindering to understanding by the app users.

### **Push Notifications**

Participants who did not complete the questionnaire daily reported that the push messages motivated them to completing the questionnaire more frequently:

*I am very pleased with the reminders... The reminders definitely encourage me to fill in the questionnaire.*

In contrast, participants who completed the questionnaire daily reported that the push messages created confusion. The messages made them question whether or not they had already completed the questionnaire on that day. They also reported the reminders to be a source of irritation when they were certain they had already completed the questionnaire that day. A potential solution offered by the participants was to create the possibility of personalizing push notifications, allowing them to be set at a user-defined time and frequency, thereby helping them to create a routine.

### **Substudy 2: Mailbox Analysis**

A total of 1080 emails were analyzed; approximately one-third of the emails were coded as irrelevant, as they comprised either duplicate mail, spam mail, mail commercials, or updates from the COVID Radar Facebook or Instagram accounts. Most emails contained questions from app users on how to interpret or answer a specific question from the self-report questionnaire (Table 1, Theme 4: content). The monitoring question about how many people came within 5 meters on a particular day resulted in extensive feedback from many users, indicating that they felt it was not reasonable to accurately estimate this value. In addition, users asked several questions about which contacts to count as “people coming within 5 meters”:

*On my way to work I pass many people that come within 5 meters of me. Should I report all these brief contacts in the Radar?*

Another monitoring question that received extensive feedback pertained to the question about the number of visitors one had received at home during a particular day. Users often questioned whether they needed to count their grandchildren aged <12 years as visitors:

*We take care of our grandchildren. Of course, they do come closer to me than 1.5 meters, but we have*

*understood that the risk of infection via children is minimal. Do we need to report our grandchildren as people we came closer than 1.5 meters?*

Several emails contained suggestions for additional questions to be included in the monitoring questionnaire. For example, many individuals suggested adding a question on vaccine status once the vaccination program started in January 2021. Another common point of feedback was that the answer format of the self-report questionnaire did not allow for any nuance. Some users reported feeling badly about their social distancing habits, for example, owing to their role as caregivers for older persons or sick family members:

*I'm sure it sometimes looks like as if I don't stick with the rules. But if the questions would allow for any nuance, especially regarding the age of visitors and informal care of, for instance, older family members, it would be a totally different outcome for me.*

Furthermore, numerous emails contained feedback on the Radar section of the app (Table 1, Theme 7: Radar section). These primarily consisted of questions on how to adequately interpret the data in the maps, as the colors in the maps did not always correspond to those in the legend. In addition, numerous emails contained comments about the answer scales of the questions regarding social distancing behavior, which were originally formatted as sliders, resulting in negative feedback from users who found the sliders difficult to use, often resulting in reporting an incorrect number of contacts. On the basis of this feedback, the sliders were replaced with a box in which the number of contacts can be entered manually, and (+) and (-) buttons can be used to manually adjust the number of contacts. This resulted in numerous emails receiving positive feedback.

### Substudy 3: Usability Testing in Individuals With Low Literacy Levels

On the basis of input from experts from the Dutch center of expertise on health disparities, and usability test sessions with individuals with low literacy levels, several optimization targets were identified regarding the design of the app. One related to improving the use of colors and color contrast to improve the visibility of letters. For example, the blue color used to color the boxes of default answers in the self-report monitoring questionnaire (Figure 1) provided limited color contrast with the white letters, sometimes leading to the default answer options being overlooked by users. Further optimization targets regarding design were to highlight the differences in the answer options by using bold text or underlining, as users with low literacy levels may otherwise have difficulty in distinguishing between slightly different answer options. A participant in one of the usability test sessions indicated the following:

*I think it says the same thing twice, I don't see what the difference is.*

Other outputs regarding design optimization were to instruct end users more explicitly on all app features. For example, that they need to scroll down for more questions, and that they need to press “submit” after completing the last question.

Regarding textual optimization, adding information was sometimes suggested to improve comprehension of questions.

For example, it was suggested we define “fever” as not everyone is aware that this is defined by a body temperature of  $\geq 38$  °C. Further, several simplifications of the text were suggested, such as to simplify or elaborate on “moderate to severe psychical exercise.” Other textual suggestions included writing out “one-and-a-half meter,” as 1.5 meters is often read as “1 to 5 meters” by individuals with low literacy.

## Discussion

### Principal Findings

A total of 3 qualitative substudies were conducted, and numerous optimization targets for the COVID Radar app were identified, both in terms of the design and content of the app. Convenience of app use, personal adjustments of the app experience, and motivation to keep using the app are key to procuring and maintaining a population of active users engaging with public health surveillance apps such as the COVID Radar. This information can be used to maximize user satisfaction, adherence, and engagement with national surveillance data collection apps.

Optimization targets as identified by the 3 studies included highlighted design elements, such as the importance of using sufficient color contrasts as well as using more visuals and less text to increase the user friendliness of the app and the comprehensibility of the corresponding texts. The results of substudy 3 highlight the importance of increasing the accessibility and comprehensibility of text for individuals with low (health) literacy levels, for example, by using language level A2 or B1 according to the Common European Framework of Reference for Languages [12] and by conducting usability tests with this target population. Limited (health) literacy levels are a prevalent and common concern worldwide [13-15]. It is essential that public health services and tools simplify their use and content to address health disparities and increase the inclusiveness of such services. High user engagement and adherence are crucial to the success of web-based monitoring tools for population-based surveillance and corresponding successful predictive modeling [3,16].

Although the feasibility and acceptability of using smartphone apps for continuous monitoring of disease seems evident from the literature [17-20], motivating individuals to use and especially keep using such tools remains challenging. The results of substudy 1 showed that the main motivations of COVID Radar users engagement were contributing to scientific research and, more specifically, helping fight the pandemic. This is in line with the available literature in the field of citizen science demonstrating that participants in citizen science projects are motivated to contribute to the scientific goals of such projects [21]. Furthermore, these results are in line with 2 important categories of community involvement in scientific projects identified by Batson et al [22]: altruism and collectivism. That is, motivation to be involved in increasing the welfare of another individual and the welfare of a group, respectively. Along similar lines, the results of substudy 1 showed that app users were eager to know what was done with their data and that in-app updates informing them about recent analyses using COVID Radar data were an important source of motivation to

keep using the app. This is in line with the results of a citizen project by Land-Zandstra et al [21], who showed that participants wanted to be kept informed about the project and its outputs. Our findings also confirm that highlighting data use can be an effective means of supporting individuals to participate in citizen science projects, as shown by Rotman et al [23].

Another potential strategy to increase user engagement and motivation is the incorporation of persuasive technology such as gamification: the use of gaming elements such as challenges and rewards to a nongame environment. Although still in its early stages, preliminary evidence supports the use of gamification to increase use and user engagement in internet interventions focusing on (mental) health and behavior change [24]. However, the effects of gamification on initiation, adherence, and engagement in a long-term monitoring population-based surveillance app remain unclear. Finally, the results from the focus group interviews demonstrated that convenience of app use (ie, user friendliness and limited time needed to complete the monitoring questionnaire) and enabling personalization of the app (ie, personal settings in terms of colors, layout, and reminders) are other important factors to consider when attempting to maximize user engagement and motivation.

### Limitations

The main limitations of substudies 1 and 2 pertain to sample bias; it is likely that the participants in the focus groups were above-mean motivated app users with more positive attitudes toward the project and the app. Likewise, email feedback is likely to be sent more frequently by highly engaged app users, whereas less motivated users or users with negative evaluations or experiences with the app may have stopped using the app and were thus not involved and reached by substudies 1 and 2. Another potential limitation is the selection bias based on the online nature of the interviews: because the focus groups were conducted using Zoom, possible participants who were not yet familiar with this program might have decided against

participation for this reason. To reduce this bias, we offered a practice round to participants who were not yet experienced in using Zoom. One participant took advantage of this opportunity. NS was available for technical assistance, which was not requested during the 3 focus group interviews. However, we cannot claim that other potential participants with a lower digital literacy were not deterred from participation. However, the participants in substudy 3 had not previously used the app; therefore, their feedback on the textual features of the app was unbiased by previous experience with the app. Furthermore, the results of substudies 1 and 3 may be limited in terms of generalizability because of the relatively low number of participants in these substudies (N=14 and N=4, respectively). The use of web-based interviewing further comprised the ability to fully assess body language. However, the interviews were conducted via videoconference, so facial expressions could be read and interpreted by the interviewer. Therefore, it is not believed that this biased our conclusions regarding the focus groups.

### Conclusions

A self-report population-based surveillance mobile app seems useful in national research on symptoms and social distancing behavior. The main motivations of citizens to use such an app are related to altruism, collectivism, and wanting to help science in developing and improving mobile self-report apps in the context of public health surveillance tools. Convenience of app use, enablement of personalization of the app, and motivational factors are key to procuring and maintaining a population of active users engaging with an app such as the COVID Radar. When designing such an app and developing content, it is important to consider its accessibility for individuals with low literacy levels. These results are not only critical to the optimization of the COVID Radar app, ultimately increasing its acceptability, inclusiveness, and adherence, but are also relevant in the broader context of citizen science projects and public health surveillance tools during the pandemic.

---

### Acknowledgments

The authors would like to thank Pharos, the Dutch center of expertise on health disparities, for conducting the expert review and usability testing of individuals with low literacy levels. The authors would also like to thank all interview participants for their valuable contributions and feedback. Funding for the project was obtained from the Netherlands Organization of Health Research and Development grant 10430042010016, 10430022010001, and 10430032010011. The funders played no role in study design, data collection and analysis, decision to publish, or manuscript preparation. The COVID Radar app was developed and maintained through a partnership between the Leiden University Medical Center and the tech company Ortec. The app is freely available in the Apple App Store and the Google Play Store.

---

### Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Final and full versions of the COVID Radar self-report questionnaire.  
[DOCX File, 20 KB - [formative\\_v6i9e36003\\_app1.docx](#) ]

---

### References

1. Total coronavirus cases in the Netherlands. Worldometers. URL: <https://www.worldometers.info/coronavirus/country/netherlands/> [accessed 2021-09-05]
2. Girum T, Lentiro K, Geremew M, Migora B, Shewamare S. Global strategies and effectiveness for COVID-19 prevention through contact tracing, screening, quarantine, and isolation: a systematic review. *Trop Med Health* 2020 Nov 23;48(1):91 [FREE Full text] [doi: [10.1186/s41182-020-00285-w](https://doi.org/10.1186/s41182-020-00285-w)] [Medline: [33292755](https://pubmed.ncbi.nlm.nih.gov/33292755/)]
3. van Dijk WJ, Saadah NH, Numans ME, Aardoom JJ, Bonten TN, Brandjes M, et al. COVID RADAR app: description and validation of population surveillance of symptoms and behavior in relation to COVID-19. *PLoS One* 2021 Jun 30;16(6):e0253566 [FREE Full text] [doi: [10.1371/journal.pone.0253566](https://doi.org/10.1371/journal.pone.0253566)] [Medline: [34191828](https://pubmed.ncbi.nlm.nih.gov/34191828/)]
4. Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, et al. COVID-19 mobile apps: a systematic review of the literature. *J Med Internet Res* 2020 Dec 09;22(12):e23170 [FREE Full text] [doi: [10.2196/23170](https://doi.org/10.2196/23170)] [Medline: [33197234](https://pubmed.ncbi.nlm.nih.gov/33197234/)]
5. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020 Jul 11;26(7):1037-1040 [FREE Full text] [doi: [10.1038/s41591-020-0916-2](https://doi.org/10.1038/s41591-020-0916-2)] [Medline: [32393804](https://pubmed.ncbi.nlm.nih.gov/32393804/)]
6. Varsavsky T, Graham MS, Canas LS, Ganesh S, Capdevila Pujol J, Sudre CH, et al. Detecting COVID-19 infection hotspots in England using large-scale self-reported data from a mobile application: a prospective, observational study. *Lancet Public Health* 2021 Jan;6(1):e21-e29. [doi: [10.1016/s2468-2667\(20\)30269-3](https://doi.org/10.1016/s2468-2667(20)30269-3)]
7. Allen WE, Altae-Tran H, Briggs J, Jin X, McGee G, Shi A, et al. Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nat Hum Behav* 2020 Sep 26;4(9):972-982. [doi: [10.1038/s41562-020-00944-2](https://doi.org/10.1038/s41562-020-00944-2)] [Medline: [32848231](https://pubmed.ncbi.nlm.nih.gov/32848231/)]
8. Bonney R, Ballard H, Jordan R, McCallie E, Phillips T, Shirk J, et al. Public participation in scientific research: defining the field and assessing its potential for informal science education. Center for Advancement of Informal Science Education (CAISE). URL: <https://files.eric.ed.gov/fulltext/ED519688.pdf> [accessed 2021-09-05]
9. Daniels N, Gillen P, Casson K, Wilson I. STEER: factors to consider when designing online focus groups using audiovisual technology in health research. *Int J Qual Method* 2019 Nov 19;18:160940691988578. [doi: [10.1177/1609406919885786](https://doi.org/10.1177/1609406919885786)]
10. Flynn R, Albrecht L, Scott SD. Two approaches to focus group data collection for qualitative health research. *Int J Qual Method* 2018 Jan 17;17(1):160940691775078. [doi: [10.1177/1609406917750781](https://doi.org/10.1177/1609406917750781)]
11. 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(1):117 [FREE Full text] [doi: [10.1186/1471-2288-13-117](https://doi.org/10.1186/1471-2288-13-117)] [Medline: [24047204](https://pubmed.ncbi.nlm.nih.gov/24047204/)]
12. Description of competence in English. Common European Framework of Reference for Languages. URL: <https://www.commoneuropenframework.org/cefr> [accessed 2021-09-05]
13. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohman LT, Rudd RR. The prevalence of limited health literacy. *J Gen Intern Med* 2005 Feb;20(2):175-184 [FREE Full text] [doi: [10.1111/j.1525-1497.2005.40245.x](https://doi.org/10.1111/j.1525-1497.2005.40245.x)] [Medline: [15836552](https://pubmed.ncbi.nlm.nih.gov/15836552/)]
14. Rajah R, Hassali M, Murugiah M. A systematic review of the prevalence of limited health literacy in Southeast Asian countries. *Public Health* 2019 Feb;167:8-15. [doi: [10.1016/j.puhe.2018.09.028](https://doi.org/10.1016/j.puhe.2018.09.028)] [Medline: [30544041](https://pubmed.ncbi.nlm.nih.gov/30544041/)]
15. Sahn LJ, Wolf MS, Curtis LM, McCarthy S. Prevalence of limited health literacy among Irish adults. *J Health Commun* 2012 Oct;17 Suppl 3(sup3):100-108. [doi: [10.1080/10810730.2012.718041](https://doi.org/10.1080/10810730.2012.718041)] [Medline: [23030564](https://pubmed.ncbi.nlm.nih.gov/23030564/)]
16. Twickler T, Hoogstraaten E, Reuwer AQ, Singels L, Stronks K, Essink-Bot ML. Low literacy and limited health skills require an answer in healthcare. *Ned Magazine Med* 2009;153:A250.
17. Aledavood T, Torous J, Triana Hoyos AM, Naslund JA, Onnela J, Keshavan M. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. *Curr Psychiatry Rep* 2019 Jun 04;21(7):49 [FREE Full text] [doi: [10.1007/s11920-019-1043-y](https://doi.org/10.1007/s11920-019-1043-y)] [Medline: [31161412](https://pubmed.ncbi.nlm.nih.gov/31161412/)]
18. Altmann S, Milsom L, Zillesen 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](https://doi.org/10.2196/19857)] [Medline: [32759102](https://pubmed.ncbi.nlm.nih.gov/32759102/)]
19. Antosik-Wójcińska AZ, Dominiak M, Chojnacka M, Kaczmarek-Majer K, Opara KR, Radziszewska W, Świącicki. Smartphone as a monitoring tool for bipolar disorder: a systematic review including data analysis, machine learning algorithms and predictive modelling. *Int J Med Inform* 2020 Jun;138:104131. [doi: [10.1016/j.ijmedinf.2020.104131](https://doi.org/10.1016/j.ijmedinf.2020.104131)] [Medline: [32305023](https://pubmed.ncbi.nlm.nih.gov/32305023/)]
20. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. *J Med Internet Res* 2017 Jul 24;19(7):e262 [FREE Full text] [doi: [10.2196/jmir.7006](https://doi.org/10.2196/jmir.7006)] [Medline: [28739561](https://pubmed.ncbi.nlm.nih.gov/28739561/)]
21. Land-Zandstra AM, Devilee JL, Snik F, Buurmeijer F, van den Broek JM. Citizen science on a smartphone: participants' motivations and learning. *Public Underst Sci* 2016 Jan 07;25(1):45-60. [doi: [10.1177/0963662515602406](https://doi.org/10.1177/0963662515602406)] [Medline: [26346340](https://pubmed.ncbi.nlm.nih.gov/26346340/)]
22. Batson CD, Ahmad N, Tsang J. Four motives for community involvement. *J Social Issues* 2002 Sep;58(3):429-445. [doi: [10.1111/1540-4560.00269](https://doi.org/10.1111/1540-4560.00269)]



23. Rotman D, Preece J, Hammock J, Procita K, Hansen D, Parr C. Dynamic changes in motivation in collaborative citizen-science projects. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12: Computer Supported Cooperative Work; Feb 11 - 15, 2012; Seattle Washington USA. [doi: [10.1145/2145204.2145238](https://doi.org/10.1145/2145204.2145238)]
24. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. *Front Psychiatry* 2016 Apr 18;7:1-7 it is article number 65 [FREE Full text] [doi: [10.3389/fpsy.2016.00065](https://doi.org/10.3389/fpsy.2016.00065)] [Medline: [27148094](https://pubmed.ncbi.nlm.nih.gov/27148094/)]

## Abbreviations

**FAQ:** frequently asked question

**LUMC:** Leiden University Medical Center

*Edited by A Mavragani; submitted 28.12.21; peer-reviewed by A Ahmed, E Metting; comments to author 15.02.22; revised version received 25.05.22; accepted 16.06.22; published 09.09.22.*

*Please cite as:*

*Splinter B, Saadah NH, Chavannes NH, Kieft-de Jong JC, Aardoom JJ*

*Optimizing the Acceptability, Adherence, and Inclusiveness of the COVID Radar Surveillance App: Qualitative Study Using Focus Groups, Thematic Content Analysis, and Usability Testing*

*JMIR Form Res 2022;6(9):e36003*

*URL: <https://formative.jmir.org/2022/9/e36003>*

*doi: [10.2196/36003](https://doi.org/10.2196/36003)*

*PMID: [35781492](https://pubmed.ncbi.nlm.nih.gov/35781492/)*

©Bas Splinter, Nicholas H Saadah, Niels H Chavannes, Jessica C Kieft-de Jong, Jiska J Aardoom. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 09.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Satisfaction With Telemedicine in Patients With Orthopedic Trauma During the COVID-19 Lockdown: Interview Study

Thomas Rauer<sup>1</sup>, MD; Julian Scherer<sup>1</sup>, MD; Pascal Stäubli<sup>2</sup>; Jonas Gerber<sup>3</sup>; Hans-Christoph Pape<sup>1</sup>, MD; Sandro-Michael Heining<sup>1</sup>, MD

<sup>1</sup>Department of Trauma Surgery, University Hospital of Zurich, Zurich, Switzerland

<sup>2</sup>Faculty of Medicine, University of Zurich, Zurich, Switzerland

<sup>3</sup>Physiotherapy Occupational Therapy, University Hospital Zurich, Zurich, Switzerland

**Corresponding Author:**

Thomas Rauer, MD

Department of Trauma Surgery

University Hospital of Zurich

Raemistrasse 100

Zurich, 8091

Switzerland

Phone: 41 0442551111

Email: [Thomas.Rauer@usz.ch](mailto:Thomas.Rauer@usz.ch)

## Abstract

**Background:** Telemedicine can take many forms, from telephone-only consultations to video consultations via a smartphone or personal computer, depending on the goals of the treatment. One of the advantages of videoconferencing is the direct visual contact between patients and therapists even over long distances. Although some telemedicine models require specially designed add-on devices, others get by with off-the-shelf equipment and software and achieve similarly successful results. This depends, among other things, on the nature of the injury, the desired outcome of therapy, and the medical consultation. In the last decade, the science and practice of telemedicine have grown exponentially and even more so during the COVID-19 pandemic. Depending on the traumatic lesion, posttraumatic and postoperative treatment and care of patients who experience trauma may require medical or physical therapy consultations in a clinic or office. However, due to the COVID-19 lockdown, direct physical follow-up was more difficult, and therefore, telemedicine solutions were sought and implemented.

**Objective:** The aim of this study was to assess satisfaction with telemedical aftercare in patients with orthopedic trauma.

**Methods:** Between March and July 2020, a standardized interview using a standardized questionnaire—Freiburg Index of Patient Satisfaction (FIPS)—among patients with orthopedic trauma who received telemedical postsurgical or physiotherapeutic care was conducted. The FIPS is composed of 5 questions regarding treatment and 1 question on the overall treatment satisfaction. Furthermore, we assessed patients' demographics and their telemedical use. Subgroup analysis was performed for age groups (<65 years vs ≥65 years), the used device, and gender.

**Results:** In total, we assessed 25 patients with a mean age of 43 (SD 24.31) years (14 female). The majority of patients (n=19, 76%) used their smartphone for consultations. The mean overall FIPS score assessed was 2.14 (SD 0.87). The mean FIPS score for younger patients was 2.23 (SD 0.90) vs 1.91 (SD 0.82) for older patients. The vast majority of the surveyed patients (n=20, 80%) were absolutely confident with their smartphone or tablet use.

**Conclusions:** Most patients surveyed stated a high satisfaction with the telemedical follow-up. Older patients showed a higher satisfaction rate than their younger counterparts. It seems that telemedical postsurgical or physiotherapeutic care is a viable option, especially in times of reduced contact, like the current COVID-19 pandemic. Thus, telemedicine offers the opportunity to ensure access to effective patient care, even over long distances, while maintaining patient satisfaction.

(*JMIR Form Res* 2022;6(9):e35718) doi:[10.2196/35718](https://doi.org/10.2196/35718)

**KEYWORDS**

COVID-19; digital; survey; telehealth; follow-up; orthopedic trauma; trauma; attitude

## Introduction

In the age of digitalization and new communication technologies, almost every generation uses smartphones or computers to be in constant exchange with the environment, to communicate with other people, or to get information anywhere and anytime [1,2]. The term telemedicine is not defined uniformly. Among other definitions, the definition of the European Commission's Health Care Telematics Program is as follows: "telemedicine is the rapid access to shared and remote medical expertise by means of telecommunication and information technologies, no matter where the patient or relevant information is located" [3]. Accordingly, a key point of telemedical treatment is the rapid exchange of information between the patient and the treating physician. The main reason for this characteristic is the widespread access to technological methods for telemedical communication, further enhanced by the independence of a geographically fixed setting for these consultations. By replacing one-on-one consultations with phone and video calls, patients can eliminate long travels to the clinic and waiting times in the waiting room before visiting the doctor [4]. The resources and costs required for this can be significantly reduced, which makes the interaction between doctor and patient much more cost-efficient [5]. Communication media used for telecommunication are phone calls, emails, videotelephony, SMS, and broadcast or telemedia [6]. Due to the COVID-19 lockdown, many hospitals had to reduce the number of their daily one-on-one consultations [7]. Therefore, they had to find a solution to ensure adequate aftercare of patients [8].

Several previous studies have shown a hypothetical acceptance (if offered by physicians) of postsurgical follow-up or surveillance via telemedical solutions as well as a high willingness to conduct video consultations in general [9,10]. Studies have shown that even physiotherapeutic interventions can be conducted effectively via videoconferencing, which is cost-effective and can give access to patients who live in remote areas [11]. However, a previous study [12] has shown that the COVID-19 crisis had no significant impact on the willingness of patients to use telemedical solutions, and data on patients' satisfaction with telemedicine, especially during national lockdowns, is lacking. Thus, the aim of this study was to determine the acceptance (subjectively) and satisfaction of telemedicine by patients who have experienced such a telemedical aftercare procedure. A standardized patient satisfaction questionnaire was used to visualize patient satisfaction in a consistent manner. Furthermore, the question of how well patients cope with digital devices, such as smartphones, computers, and tablets, was answered.

## Methods

### Patients and Setting

During the COVID-19 lockdown from March to July 2020, the department of a European Level 1 trauma center offered selected patients, for whom one-on-one therapy was not medically imperative, the option of telemedicine follow-up or physical therapy in lieu of a one-on-one consultation. Patients included in this study consented verbally to telemedical care and had the

technical requirements (ie, smartphone, computer, or phone) and expertise (subjectively) needed to attend the telemedical consultation. The telemedical follow-up or physical therapy took place either as a pure telephone consultation or as a videotelephone consultation by means of a smartphone or computer with integrated video function, where the patient could demonstrate findings such as wounds and range of motion to the treating therapist, or the therapist could instruct the patient on appropriate physiotherapeutic exercises.

Patients who medically required a one-on-one consultation, those who were declined telemedical follow-up or treatment, or could not meet the technical requirements for other reasons, such as not owning a smartphone, tablet, or computer, or not having an internet connection, were excluded.

Patients who received a telemedical follow-up or treatment were called by phone and asked to retrospectively complete a standardized patient satisfaction questionnaire (Freiburg Index of Patient Satisfaction) in relation to the treatment that had taken place after the postoperative treatment was ended. In addition, they were asked what type of device they used for telemedicine consultations and whether they are familiar with using smartphones and tablets. The questionnaire and the additional question were then completed during the telephone call as part of a standardized interview.

### Freiburg Index of Patient Satisfaction

The standardized questionnaire used in this study was the Freiburg Index of Patient Satisfaction (FIPS) questionnaire, which was developed in 2013 by Miernik et al [13] to assess treatment-related patient satisfaction. The questionnaire can be used across disciplines and regardless of the type of interventional or operative treatment. The questionnaire consists of 5 questions, called items, which are assessed using a 6-point scheme. Questions 1 to 4 can be rated on a scale from 1 (strongly agree) to 6 (strongly disagree). Question 5 can be rated on a scale from 1 (excellent) to 6 (very poor). The sum value of the points awarded is divided by the number of items after all questions have been answered, resulting in an overall score, the FIPS score. The FIPS score is thus between 1 and 6, with 1 corresponding to an excellent score and 6 to a very poor score. With the help of a regression analysis, Miernik et al [13] were able to show that neither the invasiveness of the procedure nor sociodemographic factors, such as age, gender, or school leaving certificate, have any influence on patient satisfaction. The independence of sociodemographic factors is very unusual, as they have been shown to be influencing factors in many studies. This underlines the ubiquitous applicability of the FIPS questionnaire [13-15].

The questionnaire is formulated in a way that is very understandable for everyone and can therefore be answered completely by almost all patients independently. The FIPS questionnaire is considered valid, reliable, and one-dimensional, meaning that it only focuses on subjective patient satisfaction [13]. However, patient satisfaction alone cannot provide a final verdict on the quality of a treatment or therapy, as it provides a subjective picture. Nevertheless, it is an important parameter for establishing a comprehensive quality assessment of a treatment. For the optimal validity, patient satisfaction as well

as clinical parameters or scores should be included in the overall assessment.

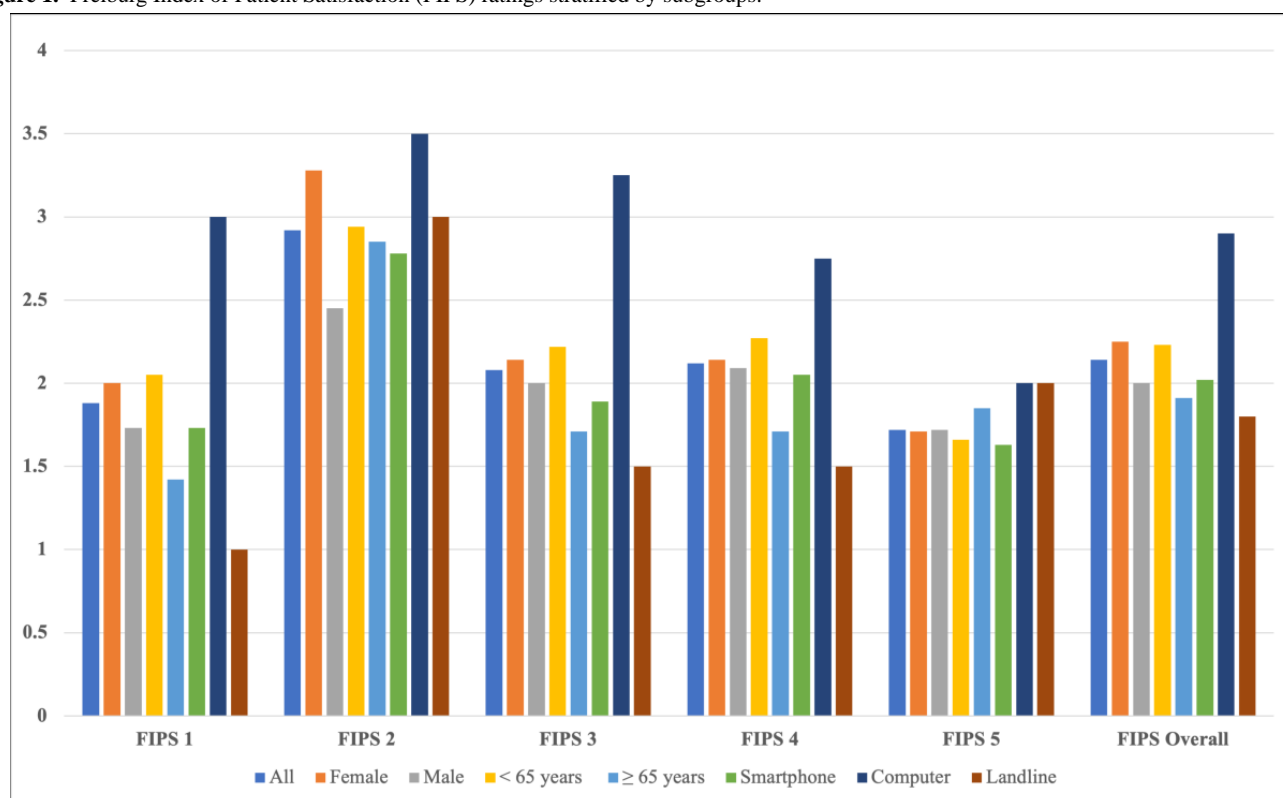
### Statistical Analysis

Further statistical analysis was performed with the use of IBM SPSS Statistics for Mac (version 26.0; IBM Corp). Data are presented as frequencies (n) and means with SDs. To assess differences in ordinal data between the groups, a nonparametric median test and a chi-square test for nominal data were used. A subgroup analysis was performed for the age groups (<65 years vs ≥65 years), gender, and the used device. The level of statistical significance was set at  $P<.05$ .

### Ethical Considerations

The local ethics committee (Kantonale Ethikkommission, Kanton Zürich) ruled that no ethical approval was necessary for this study (BASEC-Nr. Req-2020-00562).

**Figure 1.** Freiburg Index of Patient Satisfaction (FIPS) ratings stratified by subgroups.



### 1. Burden of Treatment

The mean overall score for this question was 1.88 (SD 1.23; range 1-5). There was no difference between female and male patients ( $P=.56$ ). Patients over 65 years ( $n=7$ , 28%) rated this question significantly better than their younger counterparts ( $P=.04$ ). Patients using a smartphone and landline phone had significantly better ratings than patients using a computer ( $P=.04$ ).

### 2. Recovery After Treatment

The mean overall score for this question was 2.92 (SD 1.55; range 1-6). Male patients rated this question insignificantly better than their female counterparts ( $P=.07$ ). There was no difference between the assessed age groups ( $P=.57$ ). No difference was seen between the used devices ( $P=.12$ ).

## Results

### Demographics

In total, 25 patients (14 female) with a mean age of 43 (SD 24.31; range 14-95) years were included. Of them, 7 patients were 65 years or older. Most patients ( $n=19$ , 76%) used a smartphone for the telemedical consultations, 4 (16%) used a computer, and 2 (8%) used a landline phone. There was no difference between female and male patients ( $P=.42$ ) and between age groups ( $P=.06$ ).

### FIPS Questionnaire

The results of the FIPS Questionnaire can be found in [Figure 1](#) and [Table S1](#) in [Multimedia Appendix 1](#).

### 3. Success of the Treatment

The mean overall score for this question was 2.08 (SD 1.04; range 1-5). No differences were seen in regard to gender ( $P=.84$ ) as well as between the age groups ( $P=.66$ ). Smartphone users and landline phone users showed a significantly better rating than those who used a computer ( $P=.02$ ).

### 4. Willingness to Repeat the Treatment

The mean overall score for this question was 2.12 (SD 1.30; range 1-5). No differences were seen in regard to gender ( $P=.55$ ) as well as between the age groups ( $P=.48$ ). No difference was seen between the used devices ( $P=.71$ ).

### 5. Overall Rating of the Treatment

The mean score among all participants for the overall rating was 1.72 (SD 0.79; range 1-4). No differences were seen in regard to gender ( $P=.40$ ) as well as between the age groups ( $P=.83$ ). No difference was seen between the used devices ( $P=.62$ ).

### 6. Overall FIPS

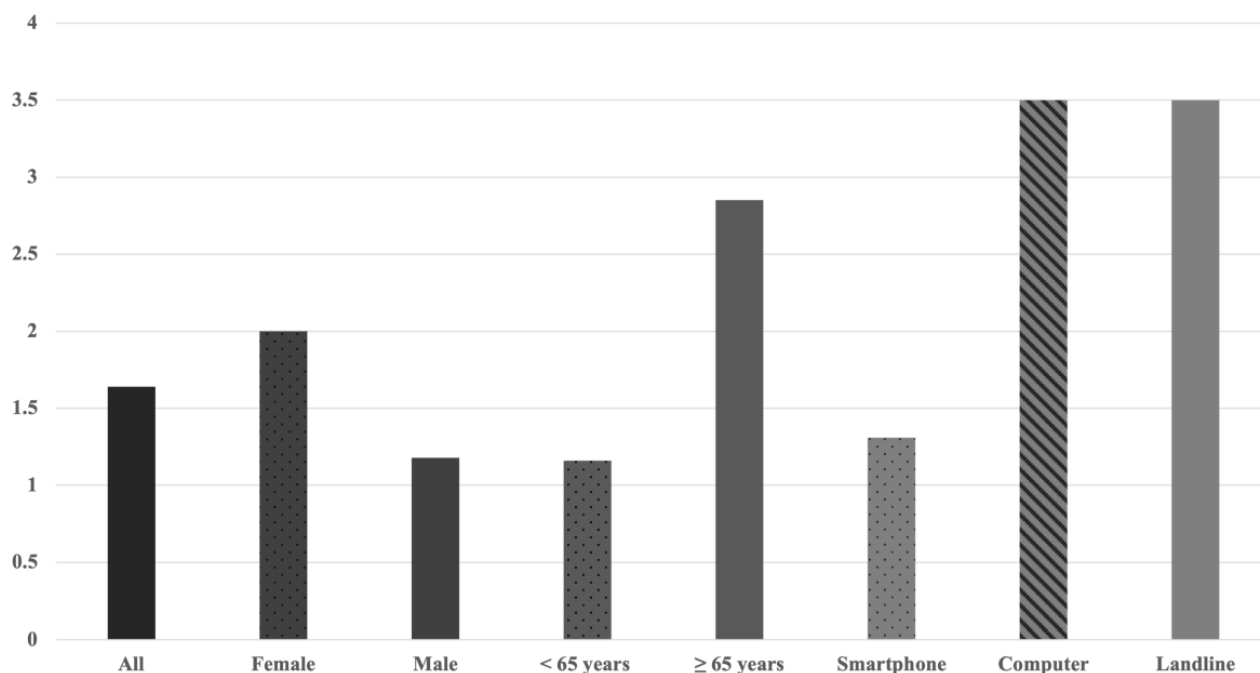
The mean overall FIPS score among all patients was 2.14 (SD 0.87; range 1-4). The mean FIPS score for younger patients was 2.23 (SD 0.90) vs 1.91 (SD 0.82) for older patients. There was no significant difference between female and male patients ( $P=.69$ ). There was also no difference between younger and

older patients ( $P=.33$ ). There was no statistical difference between the devices used ( $P=.21$ ).

### Familiarity With Smartphones and Computers or Tablets

Additionally, we found a mean score of 1.64 (SD 1.49; range 1-6) regarding electronic device familiarity among all patients. The vast majority of the surveyed patients ( $n=20$ , 80%) were absolutely confident with their smartphone or tablet use. There was no significant difference between female and male patients ( $P=.23$ ). There was also no difference between younger and older patients ( $P=.08$ ). Regarding device use, there was no significant difference between the groups ( $P=.11$ ; Figure 2).

**Figure 2.** Familiarity with smartphones and computers or tablets.



## Discussion

### Principal Findings and Comparison With Prior Work

The aim of this study was to investigate satisfaction with telemedical consultations in patients with orthopedic trauma during the COVID-19 pandemic. The answers of the 25 included patients to the FIPS questionnaire showed very positive results, reflected by an average FIPS score of 2.14 (SD 0.87). Interestingly, patients older than 65 years of age scored in the “burden of treatment” question significantly better than their younger counterparts. Furthermore, they scored the treatment in almost every item of the FIPS better than patients younger than 65 years of age. This is interesting, since other studies suggest that older patients are rather less interested in telemedicine than younger patients [5,9]. Interestingly, patients aged more than 65 years stated that they were less familiar with their electronic devices than younger patients but rated a better score for the actual telemedical treatment. The authors suggest that this indicates the easy use of telemedical devices after treatment, showing that it does not require advanced technical skills. This is all the more true for smartphones, which have become widely used in all social classes and age groups in recent

years. This could be one explanation why smartphone users and also landline phone users rated the success of the therapy significantly better than their counterparts who used a computer. No significant differences between male and female patients were found in our study. Male patients showed slightly higher confidence with their electronic devices and showed a slightly better FIPS rating than female participants. This is contrary to several studies, which showed higher use of and satisfaction with telemedicine in female patients [12]. However, a recent study on telemedical use during the COVID-19 pandemic showed findings similar to our study. Ramaswamy et al [16] assessed that younger age and female gender were associated with lower satisfaction after telemedical treatment. However, they were also able to show that during the COVID-19 pandemic, patients treated with telemedical solutions showed higher satisfaction compared to one-on-one treatments. This result supports the idea of introducing telemedical consultations with videotelephony as an alternative to one-on-one consultations. These findings as well as the results of our study suggest that telemedical solutions are valid options in terms of cost-effectiveness and travel times and especially in times of a pandemic when personal contact avoidance is desired. To our

knowledge, no such findings have been described in the current literature.

### Limitations

This study has certain limitations. It is well known that surveys have minor levels of evidence, and the outcome of this study is directly connected to participant's understanding of the questionnaire. Although there are clear trends seen in our results, these findings should be treated with caution, considering the broad age spectrum and the very diverse patient population. Patients selected for a telemedical consultation were required to have the necessary prerequisites, such as internet connection, a telemedicine-enabled digital device, and the expertise needed to operate it. This causes a certain bias in the results obtained, as patients were already selected and were sympathetic toward telemedical consultations.

As another limitation, it may be noted that we did not focus on patients with a specific injury and included patients with different patterns of injury. However, because the focus of this study was on overall patient satisfaction with telemedical treatment regardless of injury pattern, we consider the inclusion of patients with different injury patterns to have a negligible bias on the significance regarding patient satisfaction.

In addition, the limited number of included patients limits the generalizability of the statements. However, the statistically significant findings can be considered highly significant with such a small cohort.

### Conclusions

The majority of patients surveyed stated a high satisfaction with the telemedical follow-up. The results of this survey showed a positive trend in patients' attitudes toward telemedicine in both age groups, with a higher satisfaction rate in the group of older patients.

It seems that telemedical postsurgical or physiotherapeutic care is a viable option, especially in times of reduced contact, like the current COVID-19 pandemic. Our study is another component to fill the gap in the available literature on telemedicine in the treatment of patients who experience orthopedic trauma.

Further studies should include a larger number of patients and focus specifically on different trauma entities. Furthermore, a matched-pair analysis, assessing differences between telemedical aftercare and conventional aftercare, should be performed.

### Acknowledgments

All authors certify that this study received no funding to be reported.

### Conflicts of Interest

None declared.

Multimedia Appendix 1 [[DOCX File , 14 KB - formative\\_v6i9e35718\\_app1.docx](#) ]

### References

1. Individuals - mobile internet access. Eurostat. 2022. URL: [http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc\\_ci\\_im\\_i&lang=en](http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_ci_im_i&lang=en) [accessed 2022-09-02]
2. Greenwald P, Stern ME, Clark S, Sharma R. Older adults and technology: in telehealth, they may not be who you think they are. *Int J Emerg Med* 2018 Jan 03;11(1):2 [FREE Full text] [doi: [10.1186/s12245-017-0162-7](https://doi.org/10.1186/s12245-017-0162-7)] [Medline: [29299704](https://pubmed.ncbi.nlm.nih.gov/29299704/)]
3. Kailasam S, Kumar S, Dharanipragada J. Arogyasree: an enhanced grid-based approach to mobile telemedicine. *Int J Telemed Appl* 2010;2010:536237 [FREE Full text] [doi: [10.1155/2010/536237](https://doi.org/10.1155/2010/536237)] [Medline: [20467560](https://pubmed.ncbi.nlm.nih.gov/20467560/)]
4. Juhra C, Ansorg J, Back DA, John D, Kuckuck-Winkelmann A, Raschke M, et al. Online patient consultation. *Z Orthop Unfall* 2020 Aug 20;158(4):345-350. [doi: [10.1055/a-1192-7800](https://doi.org/10.1055/a-1192-7800)] [Medline: [32819007](https://pubmed.ncbi.nlm.nih.gov/32819007/)]
5. Sathiyakumar V, Apfeld J, Obremesky W, Thakore R, Sethi M. Prospective randomized controlled trial using telemedicine for follow-ups in an orthopedic trauma population: a pilot study. *J Orthop Trauma* 2015 Mar;29(3):e139-e145. [doi: [10.1097/BOT.000000000000189](https://doi.org/10.1097/BOT.000000000000189)] [Medline: [24983434](https://pubmed.ncbi.nlm.nih.gov/24983434/)]
6. Marx G, Rossaint R, Marx N. *Telemedizin: Grundlagen und Praktische Anwendung in Stationären und Ambulanten Einrichtungen*. Berlin: Springer Berlin, Heidelberg; 2021.
7. Splinter MJ, Velek P, Ikram MK, Kieboom BCT, Peeters RP, Bindels PJE, et al. Prevalence and determinants of healthcare avoidance during the COVID-19 pandemic: a population-based cross-sectional study. *PLoS Med* 2021 Nov 23;18(11):e1003854 [FREE Full text] [doi: [10.1371/journal.pmed.1003854](https://doi.org/10.1371/journal.pmed.1003854)] [Medline: [34813591](https://pubmed.ncbi.nlm.nih.gov/34813591/)]
8. Galiero R, Pafundi P, Nevola R, Rinaldi L, Acierno C, Caturano A, et al. The importance of telemedicine during COVID-19 pandemic: a focus on diabetic retinopathy. *J Diabetes Res* 2020;2020:9036847 [FREE Full text] [doi: [10.1155/2020/9036847](https://doi.org/10.1155/2020/9036847)] [Medline: [33123599](https://pubmed.ncbi.nlm.nih.gov/33123599/)]
9. Scherer J, Keller F, Pape H, Osterhoff G. Would patients undergo postoperative follow-up by using a smartphone application? *BMC Surg* 2020 Oct 07;20(1):229 [FREE Full text] [doi: [10.1186/s12893-020-00889-3](https://doi.org/10.1186/s12893-020-00889-3)] [Medline: [33028309](https://pubmed.ncbi.nlm.nih.gov/33028309/)]

10. Scherer J, Osterhoff G, Kaufmann E, Estel K, Neuhaus V, Willy C, et al. What is the acceptance of video consultations among orthopedic and trauma outpatients? A multi-center survey in 780 outpatients. *Injury* 2021 Nov;52(11):3304-3308 [FREE Full text] [doi: [10.1016/j.injury.2021.02.023](https://doi.org/10.1016/j.injury.2021.02.023)] [Medline: [33648741](https://pubmed.ncbi.nlm.nih.gov/33648741/)]
11. Mahmoud I. Usability of telemedicine in physical therapy rehabilitation: systematic review. *JMIR Rehabil Assist Technol* 2021 Jun 03;2021 [FREE Full text] [doi: [10.2196/25250](https://doi.org/10.2196/25250)] [Medline: [34287208](https://pubmed.ncbi.nlm.nih.gov/34287208/)]
12. Scherer J, Back DA, Thienemann F, Kaufmann E, Neuhaus V, Willy C, et al. The effect of COVID-19 on the willingness to use video consultations among orthopedic and trauma outpatients: a multi-center survey in 1400 outpatients. *Eur J Trauma Emerg Surg* 2022 Jun 24;48(3):2199-2206 [FREE Full text] [doi: [10.1007/s00068-021-01774-1](https://doi.org/10.1007/s00068-021-01774-1)] [Medline: [34427693](https://pubmed.ncbi.nlm.nih.gov/34427693/)]
13. Miernik A, Farin E, Kuehhas F, Karcz W, Keck T, Wengenmayer T, et al. [Freiburg index of patient satisfaction: interdisciplinary validation of a new psychometric questionnaire to describe treatment-related patient satisfaction]. *Chirurg* 2013 Jun 27;84(6):511-518. [doi: [10.1007/s00104-012-2441-4](https://doi.org/10.1007/s00104-012-2441-4)] [Medline: [23354559](https://pubmed.ncbi.nlm.nih.gov/23354559/)]
14. Khan AA, Siddiqui AZ, Mohsin SF, Mohamed BA. Sociodemographic characteristics as predictors of satisfaction in public and private dental clinics. *Pak J Med Sci* 2018 Sep 07;34(5):1152-1157 [FREE Full text] [doi: [10.12669/pjms.345.15519](https://doi.org/10.12669/pjms.345.15519)] [Medline: [30344567](https://pubmed.ncbi.nlm.nih.gov/30344567/)]
15. Fang J, Liu L, Fang P. What is the most important factor affecting patient satisfaction - a study based on gamma coefficient. *Patient Prefer Adherence* 2019 Apr;13:515-525 [FREE Full text] [doi: [10.2147/PPA.S197015](https://doi.org/10.2147/PPA.S197015)] [Medline: [31114168](https://pubmed.ncbi.nlm.nih.gov/31114168/)]
16. Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. *J Med Internet Res* 2020 Sep 09;22(9):e20786 [FREE Full text] [doi: [10.2196/20786](https://doi.org/10.2196/20786)] [Medline: [32810841](https://pubmed.ncbi.nlm.nih.gov/32810841/)]

## Abbreviations

**FIPS:** Freiburg Index of Patient Satisfaction

*Edited by A Mavragani; submitted 15.12.21; peer-reviewed by E van der Velde, N Gudi, M Bujnowska-Fedak; comments to author 22.06.22; revised version received 08.07.22; accepted 07.08.22; published 12.09.22.*

*Please cite as:*

*Rauer T, Scherer J, Stäubli P, Gerber J, Pape HC, Heining SM*

*Satisfaction With Telemedicine in Patients With Orthopedic Trauma During the COVID-19 Lockdown: Interview Study*

*JMIR Form Res* 2022;6(9):e35718

URL: <https://formative.jmir.org/2022/9/e35718>

doi: [10.2196/35718](https://doi.org/10.2196/35718)

PMID: [36040961](https://pubmed.ncbi.nlm.nih.gov/36040961/)

©Thomas Rauer, Julian Scherer, Pascal Stäubli, Jonas Gerber, Hans-Christoph Pape, Sandro-Michael Heining. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 12.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# The Impacts of Social Media Use and Online Racial Discrimination on Asian American Mental Health: Cross-sectional Survey in the United States During COVID-19

Alyan Layug<sup>1,2\*</sup>, BA; Samiksha Krishnamurthy<sup>2\*</sup>, BA; Rachel McKenzie<sup>3</sup>, MA; Bo Feng<sup>3</sup>, PhD

<sup>1</sup>Department of Asian American Studies, University of California, Davis, Davis, CA, United States

<sup>2</sup>Department of Psychology, University of California, Davis, Davis, CA, United States

<sup>3</sup>Department of Communication, University of California, Davis, Davis, CA, United States

\*these authors contributed equally

**Corresponding Author:**

Alyan Layug, BA

Department of Asian American Studies

University of California, Davis

1 Shields Ave

Davis, CA, 95616

United States

Phone: 1 530 752 1880

Email: [alayug@ucdavis.edu](mailto:alayug@ucdavis.edu)

## Abstract

**Background:** During the COVID-19 pandemic, increased social media usage has led to worsened mental health outcomes for many people. Moreover, due to the sociopolitical climate during the pandemic, the prevalence of online racial discrimination has contributed to worsening psychological well-being. With increases in anti-Asian hate, Asian and Asian American social media users may experience the negative effects of online racial discrimination in addition to the reduced psychological well-being resulting from exposure to online COVID-19 content.

**Objective:** This study aims to investigate the impact of COVID-19–related social media use and exposure to online racial discrimination during the pandemic on the mental health outcomes (ie, anxiety, depression, and secondary traumatic stress [STS]) of Asian Americans compared with those of non-Asian Americans. In addition, this study explores the mediating role of negative affect and the moderating role of racial/ethnic identification.

**Methods:** An online survey was conducted through Amazon Mechanical Turk and a university-wide research portal from March 3 to March 15, 2021. A total of 1147 participants took the survey. Participants' social media usage related to COVID-19 and exposure to 2 online forms of racial discrimination (individual and vicarious), mental health outcomes (anxiety, depression, and STS), racial/ethnic identification, negative affect, and demographics were assessed.

**Results:** Our results showed that COVID-19–related social media use, individual discrimination, and vicarious discrimination were predictors of negative mental health outcomes (anxiety, depression, and STS). Asian Americans reported higher vicarious discrimination than Latinx and White Americans, but Asian Americans' mental health outcomes did not differ substantially from those of the other racial/ethnic groups. Racial/ethnic identification moderated the relationship between both types of discrimination and STS, and negative affect served as a mediator between both types of discrimination and all 3 mental health outcomes.

**Conclusions:** These results suggest that social media exposure continues to have a dire effect on mental health during the COVID-19 pandemic. This study helps to contextualize the rise of anti-Asian American hate and its impact on mental health outcomes in the United States.

(*JMIR Form Res* 2022;6(9):e38589) doi:[10.2196/38589](https://doi.org/10.2196/38589)

**KEYWORDS**

Asian Americans; mental health; COVID-19 pandemic; racial discrimination; social media; anxiety; depression; secondary traumatic stress; negative affect; racial/ethnic identification



## Introduction

### Background

Because of a surge in social media usage by millions of individuals during the COVID-19 pandemic, social media has played an important role in the communication of disaster and health crisis-related news [1-4]. Social media may provide forms of comfort and mutual connectivity during phases of social distancing [5] and assist in sharing social, psychological, and instrumental aid [6], as well as up-to-date information regarding the pandemic [7,8]. Meanwhile, it is on social media platforms such as Twitter, Instagram, and Facebook where individuals constantly engage with disaster and health crisis-related information, such as news of deaths and mass paranoia regarding COVID-19 [9,10]. Research has shown that disaster-related content on social media during events of collective trauma (such as the COVID-19 pandemic) may negatively impact users' mental health [11-13]. Additionally, racialized hate incidents that have occurred on the internet may exacerbate the negative influence of the pandemic on racial/ethnic minority groups' mental health [14,15]. To date however, little research has conducted an integrated and contextualized empirical investigation of how the use of certain media with regard to specific types of content during significant historical and public health contexts can exert differential impacts on different groups of people, as well as the mechanisms underlying the impacts. This type of investigation is especially important, given the mixed findings of media effects documented in prior research, and can provide a nuanced understanding of the complexity of the phenomenon.

This study, therefore, aims to fill this gap. More specifically, we investigate the impact of social media usage related to COVID-19 content and social media exposure to racial discrimination on the mental health outcomes of Asian Americans (defined in this paper as anyone of Asian ethnicity living in the United States regardless of nationality) and various other racial/ethnic groups in the United States. We also examine the mediating role of negative affect and the moderating role of racial/ethnic identification in these processes.

### Social Media Usage Related to COVID-19

Social media refers to a group of internet-based applications that allow for the creation and exchange of user-generated content and enable users to participate in online social networking [16-18]. For the purposes of this study, we specifically delineated social media to include SMS text messaging, online chats, comment sections, discussion forums, online games, and social networking sites (eg, Yubo, Instagram, Twitter, and TikTok). Social media usage covers a broad range of online activities, and scholars have categorized these activities into the following two primary types: active and passive use [19-21]. Active use relates to a conscious behavior to share information [22]. This includes activities such as posting, commenting, and messaging. By contrast, passive use involves the consumption of content rather than the creation of content (eg, browsing). While scrolling or browsing content passively, social media users may have little control over what is presented to them. This is seen as "lurking" on the internet and has been

shown to be the most prominent form of usage on social media sites [23].

In this study, social media use related to COVID-19 refers to any form of social media use related to COVID-19 or the COVID-19 pandemic (eg, changes to public safety measures or a Facebook friend sharing their experience with COVID-19) that appears online, including on social media platforms. In the context of the COVID-19 pandemic, social media usage related to COVID-19 has been shown to be associated with poorer psychological outcomes [14]. Chao et al [24], for example, found that Chinese individuals' use of social media during the pandemic was positively associated with negative affect, anxiety, and stress. Social media usage predicted depression and secondary traumatic stress (STS; defined as emotional duress that results when an individual hears about the firsthand trauma experiences of another) among Wuhan residents in early February 2020, 2 weeks after Wuhan was put on lockdown [25]. A national study conducted across 30 provinces in China found that social media usage related to COVID-19 content mediated the levels of traumatic emotions among nonpatients, but did not cause negative mental health symptoms nor lead to a negative impact on mental health [26]. Furthermore, in studies on China's netizens, social media exposure to COVID-19-related content was found to be associated with poor mental health outcomes such as anxiety and depression [9]. However, to our knowledge, no current research has investigated the impact of COVID-19-related social media usage on mental health outcomes in the United States. Although Riehm et al [27] found a positive association between media exposure and mental distress in a sample of adults living in the United States during the beginning of the pandemic, they did not focus specifically on COVID-19 social media content. In accordance with the current research on social media usage and the pandemic, we predict that individuals' social media usage related to COVID-19 content would be positively associated with their anxiety, depression, and STS.

### Exposure to Online Discrimination

Exposure to online racial discrimination is an added stressor for racial/ethnic minority groups in the United States, especially during the COVID-19 pandemic [28-30]. Online racial discrimination is defined in this study as any form of discrimination denigrating or excluding individuals or groups on the basis of race through the use of symbols, voice, video, images, text, and graphic representations [31]. Online racial discrimination can be broken down into 2 forms: individual online racial discrimination and vicarious online racial discrimination, referred to as individual and vicarious discrimination, respectively, hereafter. Individual online discrimination is any form of racialized discrimination that is directly targeted at and perceived by an individual online. By contrast, vicarious online discrimination is the secondhand exposure to online racial discrimination or prejudice directed at an individual's community [32,33]. Research has shown that both individual and vicarious discrimination are significantly and positively associated with psychological distress and negative behavioral patterns such as alcohol use [34,35].

Studies have shown that various racial groups have experienced discrimination online due to their race or ethnicity [36-40], and this discrimination has continued throughout the COVID-19 pandemic [41-43]. During the COVID-19 pandemic, Asian Americans and people of Asian ethnic origin have become particularly vulnerable to racial discrimination. Asian Americans throughout US history have been and continue to be subjected to forms of xenophobia, hate, and bias [41-43], dating back to the concept of Orientalism, which refers to viewing Asians through a Western lens consisting of exoticism and a sense of superiority over the “inferior” Eastern countries. Discrimination against Asian Americans has dramatically risen in the past couple of years due to the link that has been drawn between COVID-19 and Wuhan, China, through multiple sources, especially the mass media [44]. Coinciding with this rhetoric, a large body of recent research has revealed how Asians specifically have been the subject of hate, cruelty, and anger after the outbreak of the COVID-19 pandemic [43-50]. During the start of the pandemic from March 19, 2020, to May 13, 2020, more than 1700 Anti-Asian hate incidents were reported to Stop Asian, Pacific Islander, and Mixed Asian (AAP) Hate nationwide [51].

Regardless of their pan-ethnic Asian identities (eg, South Asian, Filipino, Japanese, and Hmong), Asians have been targets of derogatory language and attacks on public social media platforms since the beginning of the pandemic [45,52]. Stop AAP Hate’s national report through June reveals that Asian Americans experienced a sharp increase in online hate from 2020 (6.1%) to 2021 (10.6%) [53]. In addition, as outlined by Cheah et al [14], Chinese Americans have experienced both individual and vicarious forms of online COVID-19–related racial discrimination. Because of the sharp increase in anti-Asian online sentiments, it can be reasonably inferred that Asian Americans may be experiencing more racial discrimination during the pandemic compared with other ethnic/racial groups in the United States. However, no research exists to analyze the rise of online discrimination of Asian Americans compared with online discrimination directed toward other racial groups. Therefore, we expect that, during the COVID-19 pandemic, individuals of Asian ethnicity would report experiencing more individual and vicarious online racial discrimination compared with individuals of non-Asian ethnicity (ie, White Americans, Black Americans) in the United States.

Prior to the COVID-19 pandemic, much research has examined the impact of racial discrimination on the mental health of ethnic minorities [38,54-64]. For example, Tynes et al [37] found that Black students and other adolescents of color were subjected to both vicarious and individual forms of online racial discrimination, which impacted students’ psychological functioning. In addition, both online vicarious and individual discrimination were significantly associated with worse psychological well-being among adults of racial/ethnic minorities (eg, Black Americans, Latinx Americans, Asian Americans) [34,35,39,40]. To our knowledge, existing research has only focused on the impacts of offline forms of racial discrimination in the context of COVID-19. Therefore, it remains unclear if online COVID-19 racial discrimination would also impact individuals’ mental health. Our study thus also

examines how the online vicarious and individual racial discrimination individuals experienced during the COVID-19 pandemic might be associated with their anxiety, depression, and STS. Online racism has impacted the mental health outcomes of various racial/ethnic groups in the United States. However, as discussed earlier, this pandemic has been a particularly difficult time for Asian Americans as anti-Asian hate has been reported to be at an all-time high. Therefore, we also aim to assess whether individuals of Asian ethnicity would report higher negative mental health outcomes (anxiety, depression, and STS) compared with individuals of non-Asian ethnicity (eg, White Americans) in the United States.

### **Mechanisms Underlying the Impact of Online Racial Discrimination on Mental Health**

To understand the mechanisms underlying the impact of online racial discrimination on individuals’ mental health during the COVID-19 pandemic, we draw upon the Differential Susceptibility to Media Effects Model (DSMM) [65]. The DSMM is a contemporary communication-based model that describes the relationships between media use and various mental health outcomes. From the perspective of the DSMM, social media use can have an adverse effect on users’ cognition and emotional state and can produce negative physiological and behavioral outcomes [65]. Besides, individual differences such as race, gender, and ethnicity can moderate the strength of social media’s effects. There are 3 response states (cognitive, executive, and excitative states) that explain the connection between the use of online media and its outcomes [65]. For our study, we investigate racial/ethnic identification as a moderator and negative affect, an excitatory response state, as a mediator.

In this study, we define racial/ethnic identification as an amalgamation of racial and ethnic identities. Racial identification refers to the “multidimensional construct that includes the strength of one’s identification with one’s racial group, a sense of attachment to other group members, an evaluation of group membership” (eg, how much the individual likes or dislikes being White, for example) and “may include group-relevant attitudes and behaviors” [66]. Ethnic identification relates to a dynamic, social construct that is reflective of cultural practices as well as the acquisition and maintenance of cultural characteristics (eg, ethnic group behaviors, knowledge and awareness of cultural beliefs, and traditions of one’s ethnic group) [67-69]. Although these are distinct social constructs, they overlap in many ways, specifically regarding how racial/ethnic identification can help can help those who have experienced racialized discrimination better understand their experience as a communal rather than a personal one [70].

Findings from past research vary in terms of whether racial/ethnic identification should be considered as a buffer [71,72] or an exacerbator of the impact of racial discrimination on negative mental health outcomes [73-75]. For example, Tynes et al’s [76] study of Black adolescents found that ethnic identification moderated the negative impact of online racial discrimination on anxiety levels such that ethnic identification lessened the impact of discrimination on mental health. Although past research has shown that ethnic/racial identification may

buffer against the negative impacts of racial discrimination [77], a recent meta-analysis revealed that the directionality of moderation varies for each ethnic/racial group [78]. Again, because past research has focused on offline forms of racial discrimination, no empirical research has examined whether the impact of online racial discrimination on individuals' mental health outcomes varies as a result of an individual's level of racial/ethnic identification. Therefore, we also examine the moderating role of racial-ethnic identification in the relationship between online racial discrimination and mental health during the COVID-19 pandemic.

Negative affect has been shown to be significantly associated with racial discrimination [79] and serves as a mediator between experiences of racial discrimination and health and psychological well-being [80,81]. For instance, Gibbons et al [81], as well as Gibbons and Stock [82], investigated the relationship between perceived racial discrimination and substance abuse, finding that negative affect mediated this relationship, such that increased perceived racial discrimination leads to greater experienced negative affect, which was positively associated with substance abuse [82,83]. Negative affect has also been shown to mediate the relationship between racial discrimination and other outcome variables such as delinquency [83] and reactive aggression for African American youth [84]. Recent research also found that social media usage concerning COVID-19 content and its effects on mental health was mediated by negative affect in a population of Chinese college students [85]. No study, however, has examined whether negative affect may mediate the impact of specific forms of online racial discrimination on psychological health. Therefore, our study also aims to assess whether negative affect can explain the effect of online racial discrimination on mental health outcomes.

## Methods

### Participants

The study was conducted through a university research participation portal as well as Amazon Mechanical Turk (MTurk) from March 3 to 15, 2021. Only individuals who were 18 years or older and were residing in the United States at the time of recruitment were qualified to participate in the study. Qualified participants recruited from the university research participation portal received course credit. Participants recruited from Amazon MTurk received a monetary compensation of US \$0.50.

### Measures

#### *Negative Affect*

The 10 - item Negative Affect Scale of the Positive and Negative Affect Schedule was used to measure participants' negative affect over the previous month [86]. Participants' experience with negative affect was measured with a 5 - point Likert - type scale (1=not really; 5=nearly every day). Participants were asked how frequently they experienced emotions/feelings such as "irritable," "distressed," and "afraid." Total scores were averaged for analyses, with higher scores on

the scale indicating greater negative affect (mean 2.26, SD 0.92; Cronbach  $\alpha$ =.925).

#### *Anxiety*

An adapted version of the 7-item Generalized Anxiety Disorder Scale was used to measure participants' anxiety during the previous month [87]. A 5-point Likert-type scale asked participants to respond on a scale from 1 (not at all) to 5 (nearly every day). Example items include "Not being able to stop or control worrying" and "Feeling afraid something awful might happen." The 7 items were averaged to indicate the level of anxiety (mean 2.55, SD 1.10; Cronbach  $\alpha$ =.932).

#### *Depression*

The 9-item Patient Health Questionnaire depression module [88] was used to measure participants' depressive symptoms during the previous month. The scale consists of 9 items measured on a 5 - point Likert - type graded response (1=not at all; 5=nearly every day). Example items include "Had little interest or pleasure in doing things" and "Had thoughts that you would be better off dead or thoughts of hurting yourself in some way." The 9 items were averaged to indicate the level of depression (mean 2.37, SD 0.99; Cronbach  $\alpha$ =.915).

#### *Secondary Traumatic Stress*

An adapted version of the 17-item Secondary Traumatic Stress Scale was used to measure participants' STS during the previous month [89]. The scale assessed participants' traumatic symptoms such as avoidance, intrusion, and arousal during the COVID-19 pandemic using 5 - point Likert - type items (1=not at all; 5=nearly every day). Example items include "It seems as if I can relive the trauma(s) experienced through the pandemic" and "I wanted to avoid thinking about the pandemic." All 17 items were averaged to indicate STS level (mean 2.30, SD 0.91; Cronbach  $\alpha$ =.939).

#### *Social Media Usage*

Social media usage was measured with 6 items adapted from the social media scale developed by Yang et al [90]. The scale measures participants' engagement in both active and passive use of social media on a scale of 1 (not at all) to 5 (regularly). Example questions include "How frequently have you paid attention to and read posts related to the COVID-19 pandemic on social media?" and "How frequently have you posted or reposted information and news related to the COVID-19 pandemic?" The 6 items were averaged to create an index of social media exposure (mean 2.33, SD 0.88; Cronbach  $\alpha$ =.859).

#### *Online Racial Discrimination*

Participants' experience of online racial discrimination during the pandemic was measured utilizing items adapted from the Online Victimization Scale [91]. A total of 5 items were adapted to measure online individual racial discrimination (eg, "I have been harassed or bothered online because of my race or ethnic group online"), and 5 items were adapted to measure online vicarious racial discrimination (eg, "People have said negative things [like rumors or name-calling] about my race or ethnic group online"). All items were measured with a 5-point Likert scale (1=never; 5=regularly) for each question. Responses to the 5 items regarding individual discrimination were averaged

to create an index of individual online racial discrimination (mean 1.47, SD 0.95; Cronbach  $\alpha$ =.962), and the overall level of vicarious racial discrimination online was calculated by averaging the 5 items for vicarious discrimination (mean 2.27, SD 1.22; Cronbach  $\alpha$ =.941).

### **Racial/Ethnic Identification**

Racial/ethnic identification was measured with an adapted version of the Asian American Identity Scale [92]. The Asian American Identity Scale was chosen due to its ability to address attachment and belonging to one's racial/ethnic group, as well as pride in and awareness of prejudice against one's racial/ethnic group. The original scale consists of 12 items, and items that referred specifically to Asian American identity or stereotypes were removed and the item "my racial/ethnic identity is important to me," was added. Participants responded using a 7-point Likert-type scale (1=strongly disagree; 7=strongly agree). Example questions include "I feel a lot of pride in the achievements of my group," "I try to carry out at least some of my group's customs and traditions (eg, relating to holidays, food, language)," and "It is important for me to learn about my group's traditions, customs, and values." Participants were asked to respond to the items based on their experiences over the last 6 months. Responses to the 7 individual items were averaged to indicate the level of racial identification (mean 4.93, SD 1.40; Cronbach  $\alpha$ =.913).

### **Covariates**

Several variables were also measured as covariates in this study. Specifically, mental health history was assessed by asking whether participants had a history of mental disorders prior to COVID-19. Participants responded either "Yes," "No," or "Prefer not to say." Demographic characteristics including age, gender, level of education, and ethnicity were also assessed. Zhao and Zhou's [85] 11-item COVID-19 stressor checklist tool was used to measure COVID-19-related stressors [85]. Participants were asked to respond "yes" or "no" to the following items: whether they experienced lockdown; known someone with confirmed or suspected (COVID-19) infection; experienced the death of a loved one(s) due to COVID-19; known an acquaintance dying from the COVID-19 infection; worked in a frontline job/as essential worker; interacted with infectious patients who had COVID-19; worked as an essential

worker during this pandemic; volunteered in epidemic prevention and control; and lacked necessities such as food, face masks, disinfectants, and medical care. The average values and standard deviations of each of the variables measured are summarized in [Multimedia Appendix 1](#).

### **Ethics Approval**

This study was approved by the institutional review board of the University of California, Davis (approval number: 1701406-1). All ethical procedures were maintained and followed, including the process of preserving web-based data privacy and security for data under the institutional review board protocol. This study abided by all applicable laws, regulations, and standard operations governing the protection of human subjects, student information, and protected health information. Participation in this study was completely voluntary. Written consent to collect data was obtained from all participants electronically and ensured that their identity would remain anonymous.

## **Results**

### **Participants**

A total of 1147 participants were included in the analysis after removing 35 incomplete responses, those who spent greater than 2 hours ( $n=11$ ) or less than 5 minutes ( $n=297$ ) on the survey, and those who incorrectly answered at least 1 of 4 attention checks ( $n=216$ ). Of the total participants ( $n=1147$ ), 676 (58.94%) were recruited through MTurk and 471 (41.06%) were undergraduates recruited from the researchers' institution. Participants ranged in age from 18 to 80 years (mean 32, SD 14 years), and most were female (701/1147, 61.12%), with 423 (36.88%) identifying as male and 15 (1.31%) identifying as nonbinary. A total of 549 (47.86%) participants reported their race as White, 306 (26.68%) as Asian/Pacific Islander or mixed Asian, 133 (11.60%) as Latinx, 109 (9.50%) as Black, 11 (0.96%) as Indigenous, and 33 (2.88%) as mixed (not Asian). Asian American and Pacific Islander participants were combined for analysis because these groups have historically been studied in tandem [93] and have been shown to experience similar increases in instances of discrimination during the COVID-19 pandemic [51,53]. Additional demographic information is provided in [Table 1](#).

**Table 1.** Participant demographics (N=1147).

Demographics	Values
Age (years), mean (SD)	32 (14.4)
<b>Gender, n (%)</b>	
Male	423 (36.88)
Female	701 (61.12)
Nonbinary	15 (1.31)
Prefer not to say	4 (0.35)
Other	4 (0.35)
<b>Race, n (%)</b>	
Black	109 (9.50)
Latinx	133 (11.60)
Indigenous	11 (0.96)
AAPI <sup>a</sup>	306 (26.68)
White	549 (47.86)
Mixed	33 (2.88)
<b>Asian ethnicity<sup>b</sup>, n (%)</b>	
Chinese	117 (10.20)
Japanese	10 (0.87)
Korean	25 (2.18)
Filipino	27 (2.35)
Vietnamese	20 (1.74)
Indian	42 (3.66)
Pakistani	14 (1.22)
Bangladeshi	3 (0.26)
Sri Lankan	2 (0.17)
Hmong	3 (0.26)
Not shown	22 (1.92)
Mixed	15 (1.31)
<b>Mechanical Turk sample (n=676)</b>	
<b>Education, n (%)</b>	
Less than high school	5 (0.44)
High school	48 (4.18)
Some college	91 (7.93)
Associates	74 (6.45)
Bachelors	304 (26.50)
Masters	137 (11.94)
Doctoral	12 (1.05)
Professional	5 (0.44)
<b>Employment, n (%)</b>	
Self-employed	91 (7.93)
Full-time	396 (34.52)
Part-time	76 (6.63)

Demographics	Values
Out of work	32 (2.79)
Not able to work	14 (1.22)
Retired	36 (3.14)
Other	31 (2.70)
<b>Income (US \$), n (%)</b>	
<10,000	27 (2.35)
10,000-20,000	52 (4.53)
20,001-40,000	137 (11.94)
40,001-60,000	146 (12.73)
60,001-80,000	122 (10.64)
80,001-100,000	70 (6.10)
100,001-120,000	55 (4.80)
>120,000	67 (5.84)
<b>College student sample (n=471)</b>	
<b>College year, n (%)</b>	
1	155 (13.51)
2	103 (8.98)
3	119 (10.37)
4	80 (6.97)
5	14 (1.22)
<b>Major, n (%)</b>	
Social science	246 (21.45)
Humanities	21 (1.83)
Math/engineering	22 (1.92)
Biological science	125 (10.90)
Physical science	7 (0.61)
Undeclared	18 (1.57)
Professional school	1 (0.09)

<sup>a</sup>AAPI: Asian American, Pacific Islander, and Mixed Race Asian identities (see [Multimedia Appendix 2](#) for a more in-depth summary of the Asian ethnic subgroups).

<sup>b</sup>A total of 300 participants noted their Asian ethnicity.

## Predictors of Mental Health Outcomes

Multiple hierarchical regression models were run utilizing SPSS version 27 (IBM Corp) to assess social media usage and individual and vicarious discrimination as predictors of mental health outcomes. Step 1 included the covariates of mental health history, the experience of COVID-19 stressors, age, gender, race, education, and income. The independent variables of interest (COVID-19–related social media use, individual online discrimination, and vicarious online discrimination) were added in step 2. Full results from the hierarchical regressions are presented in [Multimedia Appendices 3-5](#). The results showed that COVID-19–related social media use was positively associated with STS ( $\beta=.39$ ,  $P<.001$ ), depression ( $\beta=.36$ ,  $P<.001$ ), and anxiety ( $\beta=.30$ ,  $P<.001$ ). Individual online perceived discrimination was a significant positive predictor of

STS ( $\beta=.52$ ,  $P<.001$ ), depression ( $\beta=.53$ ,  $P<.001$ ), and anxiety ( $\beta=.41$ ,  $P<.001$ ). Similarly, vicarious online perceived discrimination was positively associated with STS ( $\beta=.39$ ,  $P<.001$ ), depression ( $\beta=.39$ ,  $P<.001$ ), and anxiety ( $\beta=.33$ ,  $P<.001$ ).

## Racial/Ethnic Differences

To test whether Asian participants differed from other racial/ethnic groups in reports of individual and vicarious online perceived discrimination and negative mental health outcomes, a 1-way ANOVA and Bonferroni post hoc test were utilized. Black, Latinx, White, and AAPI groups were compared in the analysis. Because of the small sample size, Indigenous ( $n=11$ ) and mixed ( $n=33$ ) racial/ethnic groups were not included. There was a statistical difference between racial/ethnic groups for individual ( $F_{3,1083}=31.23$ ,  $P<.001$ ;  $\eta^2=0.080$ ) and vicarious

discrimination ( $F_{3,1085}=46.40$ ,  $P<.001$ ;  $\eta^2=0.114$ ). Black participants (mean 2.21, SD 1.31) reported significantly higher individual discrimination than Asian (mean 1.34, SD 0.79;  $P<.001$ ), Latinx (mean 1.13, SD 0.43;  $P<.001$ ), and White (mean 1.47, SD 0.97;  $P<.001$ ) participants. Additionally, Black participants (mean 2.98, SD 1.18;  $P<.001$ ) reported significantly higher vicarious discrimination than Latinx (mean 2.24, SD 0.97;  $P<.001$ ) and White (mean 1.90, SD 1.14;  $P<.001$ ) participants. Asian participants reported significantly higher vicarious discrimination (mean 2.69, SD 1.23) than Latinx (mean 2.23, SD 0.97;  $P<.001$ ) and White participants (mean 1.90, SD 1.14;  $P<.001$ ). Therefore, partial support for Asian Americans reporting higher discrimination was found.

There was a statistical difference between racial/ethnic groups for depression ( $F_{3,1068}=6.73$ ,  $P<.001$ ;  $\eta^2=0.019$ ) and anxiety ( $F_{3,1076}=3.40$ ,  $P=.017$ ;  $\eta^2=.009$ ) but not for STS. Post hoc analyses showed that AAPI participants did not differ significantly in STS from White ( $P>.99$ ), Black ( $P>.99$ ), or Latinx ( $P=.97$ ) participants. AAPI participants also did not differ significantly in depression from White ( $P=.05$ ), Black ( $P>.99$ ), or Latinx ( $P=.23$ ) participants or differ significantly in anxiety from White ( $P=.42$ ), Black ( $P>.99$ ), or Latinx ( $P=.80$ ) participants.

Considering that Chinese participants made up the largest ethnic subgroup in the sample of Asian participants, an unpaired independent sample *t* test (2-tailed) was used to determine whether Chinese participants experienced different levels of discrimination and mental health outcomes compared with other Asian ethnicities. Chinese participants experienced lower levels of STS ( $n=114$ ; mean 1.99, SD 0.75) compared with other Asian participants ( $n=185$ ; mean 2.48, SD 0.91;  $t_{274.78}=-5.016$ ,  $P<.001$ ). Chinese participants also experienced lower levels of depression ( $n=114$ ; mean 2.18, SD 0.87) compared with other Asian participants ( $n=187$ ; mean 2.61, SD 0.99;  $t_{262.63}=-3.991$ ,  $P<.001$ ) and anxiety ( $n=115$ ; mean 2.30, SD 0.98) compared with other Asian participants ( $n=188$ ; mean 2.61, SD 0.99;  $t_{271.828}=-3.978$ ,  $P<.001$ ). Lastly, other Asian participants ( $n=187$ ; mean 2.84, SD 1.24) experienced more vicarious discrimination compared with Chinese participants ( $n=117$ ; mean 2.46, SD 1.19;  $t_{253.02}=-2.692$ ,  $P=.008$ ).

### Racial/Ethnic Identification

To address the moderating role of racial/ethnic identification in the relationship between individual and vicarious online racial discrimination and the 3 mental health outcomes, the interaction between individual and vicarious online racial discrimination and racial/ethnic identification was tested utilizing model 1 of the PROCESS macro for SPSS by Hayes [94]. Racial/ethnic identification was a significant moderator between both individual ( $\beta=.07$ ,  $P=.01$ ) and vicarious ( $\beta=.06$ ,  $P=.002$ ) online racial discrimination and STS such that the association between discrimination and STS was stronger as racial identity increased. This was not the case for depression or anxiety for either type of discrimination.

### Negative Affect

Lastly, the role of negative affect as a mediator in the relationship between individual and vicarious discrimination and the 3 mental health outcomes was assessed using model 4 of the PROCESS macro. Negative affect mediated the effects of individual discrimination on STS (indirect effect [IE] 0.3375, 95% CI 0.27-0.42) and depression (IE 0.3332, 95% CI 0.26-0.41), and fully mediated the effects of individual discrimination on anxiety (IE 0.3974, 95% CI 0.31-0.49). Negative affect also mediated the effects of vicarious discrimination on STS (IE 0.2379, 95% CI 0.18-0.31) and depression (IE 0.2369, 95% CI 0.18-0.30) and fully mediated the effects of individual discrimination on anxiety (IE 0.2711, 95% CI 0.21-0.35).

## Discussion

### Principal Findings

The main goal of this study was to examine the impact of social media use and racial discrimination on individuals' mental health (ie, symptoms of depression, anxiety, and STS) among different ethnic and racial groups in the United States during the COVID-19 pandemic. In addition, we investigated the moderating effects of racial/ethnic identification and the mediating role of negative affect in those relationships.

The findings of our study support several general conclusions. First, as expected, we found that increased use of social media during the public health crisis of the COVID-19 pandemic had a negative impact on users' mental health. This finding resonates with those from similar studies that have examined social media usage and its impact on people's mental health during the pandemic in other countries such as China [9,24,95], Bolivia [96], and Bangladesh [97].

This finding holds important pragmatic implications. Although media can provide social connectivity and public health updates, it can also contribute to the negative mental well-being of users. Social media has fueled the rapid spread of misinformation (eg, false news and racial discrimination) during this pandemic [98-101], which may create a sense of panic and confusion, leading to problematic social media usage (PSMU)—a form of addiction to social media that is associated with anxiety, depressive symptoms, and low self-esteem [102]. There is evidence that PSMU has increased during the pandemic, that is, when social isolation became the new norm [103-106]. Although this study did not directly examine PSMU, it is likely that the observed negative association between social media use specifically related to COVID-19 content and mental health outcomes can be partially attributed to PSMU. Future research should explore the relationship between PSMU and COVID-19-related social media use and their impact on mental health.

The connection between social media usage related to the COVID-19 pandemic and negative mental health outcomes shows the importance not only of further research into the topic, but also the need for actions such as policy change and educational campaigns [9,107]. In particular, concrete recommendations may be introduced and encouraged to combat

misinformation and PSMU. For instance, social media platforms can reinforce more website policing to delineate false online information and regulate the content. With negative information being regulated, this may mitigate the negative media effects on users. Users may also seek fact-checking labels on posts as well as other sources of news to verify online statements. Additionally, national policy may be implemented to further limit and moderate social media content that includes hateful speech and discriminatory online behaviors. In addition, social media users may want to reduce their screen time and actively become mindful consumers of media. Online users can use various methods such as apps to track their screen time or disabling notifications to limit usage. Although placing the onus on individual social media users is not the only approach to reducing the negative impacts of social media use, recent studies have demonstrated that behavioral, self-imposed regulations may be powerful methods to regulate usage and promote healthy online habits. Additionally, future long-term strategies have included the promotion of digital health guidelines that are taught in public sectors and enforced by evidence-based policy and legislation [101].

Consistent with our predictions, individual and vicarious forms of online discrimination experienced during the COVID-19 pandemic were found to be positively associated with STS, depression, and anxiety. Our data also revealed several interesting findings regarding different ethnic groups' experiences during the pandemic. On the one hand, Asian participants in our study reported significantly higher vicarious discrimination than both Latinx and White participants. This finding corroborates recent research concerning discrimination experienced by Asian Americans [53,73,90]. In relation to our large Chinese subpopulation, we also found that Chinese participants experienced lower levels of each of the mental health outcomes and lower levels of vicarious discrimination compared with other Asian participants, which reflects the rise in online hate toward the group of Asian Americans as a whole [53]. On the other hand, Asian American participants in our study also did not perceive significantly more individual online discrimination compared with the other racial/ethnic groups. Although this study did not compare sources of discrimination, previous research suggests that this finding may be related to the fact that multiple racial/ethnic minority communities have experienced distinct forms of discrimination during this pandemic in the spheres of health care, accessibility to resources, and vaccines, and more [108-111]. Research on Black Americans, for example, points to structural health discrimination and bioethical concerns causing the "epidemic" of premature black death [112-114]. Additionally, there is evidence showing that Indigenous and Latinx groups had the highest hospitalization and mortality rates in the United States during the pandemic [115-119].

Contrary to our expectations, in comparison to other large racial groups in our study, Black Americans reported experiencing more online discrimination compared with Asian Americans. In particular, we found that Black Americans reported the highest levels of individual as well as vicarious discrimination among all the racial/ethnic groups examined in this study, corresponding to the vast prevalence of online hatred directed

toward Black Americans throughout US history [60,120-122]. These findings may be related to the collective hardships experienced by Black Americans, who continued to experience heightened forms of hate and systemic oppression made salient during the COVID-19 pandemic through media coverage and social movements such as Black Lives Matter [123-125].

Black Americans also reported the worst mental health outcomes. As a result, this study emphasizes the importance of recognizing the mental health struggles of this community, especially due to the disproportionate number of deaths in the Black community from COVID-19 health disparities [126]. Additionally, racism and bias do impact medical care, as research shows that Black women are much more likely to not receive appropriate diagnoses for depression, and are more likely to be seen as having psychosis than depression [127].

The findings from our study also revealed that racial/ethnic identification moderates both individual and vicarious discrimination for STS. In other words, it can be understood that the STS levels of individuals who identify more strongly with their racial/ethnic community are more likely to be impacted by their experience with individual and vicarious discrimination. If one closely aligns themselves and takes pride in their racial/ethnic community, attacks on that identity or on others who share that identity would be more personally devastating to them than to someone who is not as aligned with their racial/ethnic community. Considering that the extent to which one identifies with their racial/ethnic identity can be a powerful indicator of how adversely racial discrimination can impact their mental health outcomes, certain individuals may benefit more from reducing their exposure to online content or taking steps toward limiting their mental health effects of social media use by seeking mental health care. Racial/ethnic identification was not a moderator for the other 2 mental health outcomes. Taken as a whole, our findings regarding online racial discrimination suggest the necessity for some policy and educational interventions. For example, if policy-level regulations, such as content blocking, were implemented on a national level, there is a possibility that racial hate online could significantly decrease. Moreover, culturally sensitive educational plans for children and youth can be made to promote cross-cultural understanding and social awareness of social issues online. These potential educational opportunities that specifically discuss online racial discrimination and hate could be useful for reducing the prevalence of hateful speech on social media as well as providing social media users with ways to reduce the negative mental health impacts of online racial discrimination.

### Comparison With Prior Work

Our study connects with prior research in several ways. First, regarding the impact of social media use and discrimination on mental health, our findings are consistent with past research that shows that increases in media use are common during experiences of collective trauma [128,129]. Importantly, it echoes past findings regarding the impact of media usage related to trauma and disaster content on psychological well-being [95,130-136]. Additionally, our finding of increased racial discrimination being correlated with worse negative mental



health outcomes is also in line with previous work [31,54-63]. Although our study did not distinguish between online racism and online discrimination, it aligns with recent studies, demonstrating that online discrimination, including online racism, during the pandemic could have a substantial negative impact on individuals' mental health [34,35,38,39]. Second, we were intrigued to find that White Americans in our study reported experiencing more individual discrimination than Asian Americans during the COVID-19 pandemic. Previous research has shown that White Americans perceive different types of discrimination compared to those experienced by people of color and that White Americans report experiencing discrimination as a racial group [137-139]. This anti-White bias is empirically shown to be linked to worse mental wellness for White participants [137]. National surveys have indicated that White individuals have reported discrimination in several public sectors [138,139]. For instance, a national poll conducted by Harvard found that 19% of White Americans in its sample said that they have been personally discriminated against because they are White when applying for jobs, and 28% of White individuals who believe in the existence of anti-White discrimination reported that they have personally experienced hearing insensitive comments related to their race [139].

Finally, our findings suggest that negative affect could aid in explaining the relationship between online racial discrimination and mental health outcomes. In the past, online and offline experiences with racial discrimination have been shown to trigger certain negative emotional or physical responses (eg, stress), which in turn may lead to worse mental well-being. However, this was not explicitly examined as negative affect [36,140,141]. Our findings suggest that negative affect may explain how racial discrimination is linked to worsening behavioral [85] and psychological well-being [82,83] in offline and online contexts. This finding is also consistent with the pathway proposed by the DSMM, that is, negative emotions or affective states may explain the relationship between online racial discrimination on social media and negative mental health outcomes.

### Limitations and Recommendations for Future Research

A primary limitation of our study was the method used for data collection, specifically the use of college students as well as participants recruited from Amazon MTurk. Although the college student sample primarily identified as Asian or mixed Asian (232/471, 49.3%), other racial/ethnic groups, such as Black individuals (109/1147, 9.50%), were underrepresented in our sample; thus, experiences of discrimination online may have been underreported. Additionally, some studies indicate that MTurk samples may not be as representative as national probability samples [99,142,143]. Considering that we were

unable to conduct a pan-ethnic survey and that the majority of the participants in our survey were White, it is reasonable to assume that if we opted for a different platform for data collection, our results may have been different and may have better represented the experience of all Americans [83]. Because of our limited sample size, we could not conduct a national survey. Our cross-sectional study was not able to analyze and compare all experiences of individuals from all race/ethnic groups in the US population.

Another limitation of our study was the time frame for which we kept the survey open. At the time of data collection, the United States was approaching an end to strict COVID-19 guidelines, and data collection was completed in 2 weeks in consideration of the rapidly evolving nature of the pandemic. However, the time frame during which we collected our data was not during the peak of the pandemic, nor the peak of racialized discrimination against Asian Americans. As such, we believe that the true salience of discrimination against AAPI was not fully captured by this survey, as many hate crimes were reported taking place in 2020 compared with the time of data collection [53].

Because of the limited size and representativeness of our sample, we were unable to run pan-ethnic comparisons of various Asian ethnicities (eg, comparing Japanese Americans with Indian Americans). Future research should conduct an examination of pan-ethnic comparisons of various Asian ethnicities, which will provide more insights into the experiences of Asian American communities. As previously mentioned, due to our data collection methods and inability to keep our survey available for longer periods of data collection, we could not examine changes in perceived online discrimination or the longer effects of exposure to online discrimination. Therefore, it remains important for future research to capture how changes in the prevalence of online discrimination in the wake of global events impact racial/ethnic minority groups.

### Conclusions

Our study demonstrated a high prevalence of mental health concerns associated with social media usage related to COVID-19, as well as a link between perceived online discrimination and poor mental health during the pandemic. Our results concerning online social media may reflect the offline experiences and struggles of several racial/ethnic communities. Our study used a self-report survey data collection method to examine Asian Americans and various racial/ethnic groups living in the United States as they experienced perceived discrimination during the pandemic. This research and additional studies should broaden the scientific community's and the public's understanding of different ethnic groups' social media use and its impact on their mental health, especially during a period of crisis.

### Acknowledgments

We thank the Department of Communication, University of California, Davis for funding.

## Conflicts of Interest

None declared.

### Multimedia Appendix 1

Summary of independent and dependent variables.

[DOCX File , 12 KB - [formative\\_v6i9e38589\\_app1.docx](#) ]

### Multimedia Appendix 2

Summary of Asian ethnicities.

[DOCX File , 9 KB - [formative\\_v6i9e38589\\_app2.docx](#) ]

### Multimedia Appendix 3

Ordinal logistic regression results for secondary traumatic stress.

[DOCX File , 14 KB - [formative\\_v6i9e38589\\_app3.docx](#) ]

### Multimedia Appendix 4

Ordinal logistic regression results for depression.

[DOCX File , 13 KB - [formative\\_v6i9e38589\\_app4.docx](#) ]

### Multimedia Appendix 5

Ordinal logistic regression results for anxiety.

[DOCX File , 13 KB - [formative\\_v6i9e38589\\_app5.docx](#) ]

## References

1. Taylor D. COVID-19: Social media use goes up as country stays indoors. Victoria News. 2020 Mar 31. URL: <https://www.vicnews.com/news/covid-19-social-media-use-goes-up-as-country-stays-indoors/> [accessed 2022-09-08]
2. Statista Research Department. Topic: Social media use during coronavirus (COVID-19) worldwide. Statista. 2021 Mar 19. URL: <https://www.statista.com/topics/7863/social-media-use-during-coronavirus-covid-19-worldwide/> [accessed 2022-09-08]
3. 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](https://doi.org/10.1016/j.ijinfomgt.2020.102171)] [Medline: [32836633](https://pubmed.ncbi.nlm.nih.gov/32836633/)]
4. Koeze E, Popper N. The virus changed the way we internet. The New York Times. 2020 Apr 7. URL: <https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html> [accessed 2022-09-08]
5. Cauberghe V, Van Wesenbeeck I, De Jans S, Hudders L, Ponnet K. How adolescents use social media to cope with feelings of loneliness and anxiety during COVID-19 lockdown. Cyberpsychol Behav Soc Netw 2021 Apr 20;24(4):250-257. [doi: [10.1089/cyber.2020.0478](https://doi.org/10.1089/cyber.2020.0478)] [Medline: [33185488](https://pubmed.ncbi.nlm.nih.gov/33185488/)]
6. Merchant RM, Lurie N. Social media and emergency preparedness in response to novel coronavirus. JAMA 2020 May 26;323(20):2011-2012. [doi: [10.1001/jama.2020.4469](https://doi.org/10.1001/jama.2020.4469)] [Medline: [32202611](https://pubmed.ncbi.nlm.nih.gov/32202611/)]
7. Pollett S, Rivers C. Social media and the new world of scientific communication during the COVID-19 pandemic. Clin Infect Dis 2020 Nov 19;71(16):2184-2186 [FREE Full text] [doi: [10.1093/cid/ciaa553](https://doi.org/10.1093/cid/ciaa553)] [Medline: [32396623](https://pubmed.ncbi.nlm.nih.gov/32396623/)]
8. Peyravi M, Ahmadi Marzaleh M. The Double-Edged Sword of Iranian Social Media Against COVID-19. Prehosp. Disaster med 2020 May 21;35(4):468-469. [doi: [10.1017/s1049023x20000722](https://doi.org/10.1017/s1049023x20000722)]
9. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020 Apr 16;15(4):e0231924 [FREE Full text] [doi: [10.1371/journal.pone.0231924](https://doi.org/10.1371/journal.pone.0231924)] [Medline: [32298385](https://pubmed.ncbi.nlm.nih.gov/32298385/)]
10. Banerjee D, Meena K. COVID-19 as an "Infodemic" in Public Health: Critical Role of the Social Media. Front Public Health 2021;9:610623 [FREE Full text] [doi: [10.3389/fpubh.2021.610623](https://doi.org/10.3389/fpubh.2021.610623)] [Medline: [33816415](https://pubmed.ncbi.nlm.nih.gov/33816415/)]
11. Pfefferbaum B, Newman E, Nelson SD, Nitiéma P, Pfefferbaum RL, Rahman A. Disaster media coverage and psychological outcomes: descriptive findings in the extant research. Curr Psychiatry Rep 2014 Sep 27;16(9):464 [FREE Full text] [doi: [10.1007/s11920-014-0464-x](https://doi.org/10.1007/s11920-014-0464-x)] [Medline: [25064691](https://pubmed.ncbi.nlm.nih.gov/25064691/)]
12. Van Bortel T, Basnayake A, Wurie F, Jambai M, Koroma AS, Muana AT, et al. Psychosocial effects of an Ebola outbreak at individual, community and international levels. Bull World Health Organ 2016 Mar 01;94(3):210-214 [FREE Full text] [doi: [10.2471/BLT.15.158543](https://doi.org/10.2471/BLT.15.158543)] [Medline: [26966332](https://pubmed.ncbi.nlm.nih.gov/26966332/)]
13. 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](https://doi.org/10.2196/19556)] [Medline: [32369026](https://pubmed.ncbi.nlm.nih.gov/32369026/)]
14. Cheah C, Wang C, Ren H, Zong X, Cho H, Xue X. COVID-19 Racism and Mental Health in Chinese American Families. Pediatrics 2020 Nov;146(5):e2020021816. [doi: [10.1542/peds.2020-021816](https://doi.org/10.1542/peds.2020-021816)] [Medline: [32873719](https://pubmed.ncbi.nlm.nih.gov/32873719/)]

15. Woo B, Jun J. COVID-19 Racial Discrimination and Depressive Symptoms among Asians Americans: Does Communication about the Incident Matter? *J Immigr Minor Health* 2022 Feb 06;24(1):78-85 [FREE Full text] [doi: [10.1007/s10903-021-01167-x](https://doi.org/10.1007/s10903-021-01167-x)] [Medline: [33677775](https://pubmed.ncbi.nlm.nih.gov/33677775/)]
16. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* 2010 Jan;53(1):59-68. [doi: [10.1016/j.bushor.2009.09.003](https://doi.org/10.1016/j.bushor.2009.09.003)]
17. Brooks S. Does personal social media usage affect efficiency and well-being? *Computers in Human Behavior* 2015 May;46:26-37. [doi: [10.1016/j.chb.2014.12.053](https://doi.org/10.1016/j.chb.2014.12.053)]
18. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using Social Media as a Research Recruitment Tool: Ethical Issues and Recommendations. *Am J Bioeth* 2017 Mar;17(3):3-14 [FREE Full text] [doi: [10.1080/15265161.2016.1276644](https://doi.org/10.1080/15265161.2016.1276644)] [Medline: [28207365](https://pubmed.ncbi.nlm.nih.gov/28207365/)]
19. Frison E, Eggermont S. Exploring the Relationships Between Different Types of Facebook Use, Perceived Online Social Support, and Adolescents' Depressed Mood. *Social Science Computer Review* 2015 Jan 21;34(2):153-171. [doi: [10.1177/0894439314567449](https://doi.org/10.1177/0894439314567449)]
20. Verduyn P, Lee DS, Park J, Shablack H, Orvell A, Bayer J, et al. Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *J Exp Psychol Gen* 2015 Apr;144(2):480-488. [doi: [10.1037/xge0000057](https://doi.org/10.1037/xge0000057)] [Medline: [25706656](https://pubmed.ncbi.nlm.nih.gov/25706656/)]
21. Yang C. Instagram Use, Loneliness, and Social Comparison Orientation: Interact and Browse on Social Media, But Don't Compare. *Cyberpsychol Behav Soc Netw* 2016 Dec;19(12):703-708. [doi: [10.1089/cyber.2016.0201](https://doi.org/10.1089/cyber.2016.0201)] [Medline: [27855266](https://pubmed.ncbi.nlm.nih.gov/27855266/)]
22. Pagani M, Hofacker CF, Goldsmith RE. The influence of personality on active and passive use of social networking sites. *Psychol. Mark* 2011 Apr 01;28(5):441-456. [doi: [10.1002/mar.20395](https://doi.org/10.1002/mar.20395)]
23. Krasnova H, Wenninger H, Widjaja T, Buxmann P. Envy on Facebook: a Hidden Threat to Users' Life Satisfaction? In: *Wirtschaftsinformatik Proceedings 2013*. 2013 Jan 1 Presented at: *Wirtschaftsinformatik 2013*; Month dd yyy; Location URL: <https://aisel.aisnet.org/wi2013/92>
24. 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](https://doi.org/10.1016/j.janxdis.2020.102248)] [Medline: [32505918](https://pubmed.ncbi.nlm.nih.gov/32505918/)]
25. Zhong B, Huang Y, Liu Q. Mental health toll from the coronavirus: Social media usage reveals Wuhan residents' depression and secondary trauma in the COVID-19 outbreak. *Comput Human Behav* 2021 Jan;114:106524 [FREE Full text] [doi: [10.1016/j.chb.2020.106524](https://doi.org/10.1016/j.chb.2020.106524)] [Medline: [32836728](https://pubmed.ncbi.nlm.nih.gov/32836728/)]
26. Zhong B, Jiang Z, Xie W, Qin X. Association of Social Media Use With Mental Health Conditions of Nonpatients During the COVID-19 Outbreak: Insights from a National Survey Study. *J Med Internet Res* 2020 Dec 31;22(12):e23696 [FREE Full text] [doi: [10.2196/23696](https://doi.org/10.2196/23696)] [Medline: [33302256](https://pubmed.ncbi.nlm.nih.gov/33302256/)]
27. Riehm KE, Hologue C, Kalb LG, Bennett D, Kapteyn A, Jiang Q, et al. Associations Between Media Exposure and Mental Distress Among U.S. Adults at the Beginning of the COVID-19 Pandemic. *Am J Prev Med* 2020 Nov;59(5):630-638 [FREE Full text] [doi: [10.1016/j.amepre.2020.06.008](https://doi.org/10.1016/j.amepre.2020.06.008)] [Medline: [33011008](https://pubmed.ncbi.nlm.nih.gov/33011008/)]
28. Litam SDA. "Take Your Kung-Flu Back to Wuhan": Counseling Asians, Asian Americans, and Pacific Islanders With Race-Based Trauma Related to COVID-19. *TPC* 2020 Jun;10(2):144-156. [doi: [10.15241/sdal.10.2.144](https://doi.org/10.15241/sdal.10.2.144)]
29. Tavernise Jr SR. Spit On, Yelled At, Attacked: Chinese-Americans Fear for Their Safety. *The New York Times*. 2020 Mar 23. URL: <https://www.nytimes.com/2020/03/23/us/chinese-coronavirus-racist-attacks.html> [accessed 2022-09-08]
30. Lee S, Waters SF. Asians and Asian Americans' experiences of racial discrimination during the COVID-19 pandemic: Impacts on health outcomes and the buffering role of social support. *Stigma and Health* 2021 Feb;6(1):70-78. [doi: [10.1037/sah0000275](https://doi.org/10.1037/sah0000275)]
31. Tynes B, Rose C, Hiss S, Umaña-Taylor AJ, Mitchell K, Williams D. Virtual Environments, Online Racial Discrimination, and Adjustment among a Diverse, School-Based Sample of Adolescents. *Int J Gaming Comput Mediat Simul* 2016;6(3):1-16 [FREE Full text] [doi: [10.4018/ijgcms.2014070101](https://doi.org/10.4018/ijgcms.2014070101)] [Medline: [27134698](https://pubmed.ncbi.nlm.nih.gov/27134698/)]
32. Tynes B. Online Racial discrimination: a Growing Problem for Adolescents. American Psychological Association. Washington, DC: American Psychological Association; 2015. URL: <https://www.apa.org/science/about/psa/2015/12/online-racial-discrimination> [accessed 2022-09-09]
33. Heard-Garris N, Cale M, Camaj L, Hamati M, Dominguez T. Transmitting Trauma: A systematic review of vicarious racism and child health. *Soc Sci Med* 2018 Feb;199:230-240. [doi: [10.1016/j.socscimed.2017.04.018](https://doi.org/10.1016/j.socscimed.2017.04.018)] [Medline: [28456418](https://pubmed.ncbi.nlm.nih.gov/28456418/)]
34. Keum BT, Ahn LH. Impact of online racism on psychological distress and alcohol use severity: Testing ethnic-racial socialization and silence about race as moderators. *Computers in Human Behavior* 2021 Jul;120:106773. [doi: [10.1016/j.chb.2021.106773](https://doi.org/10.1016/j.chb.2021.106773)]
35. Keum BTH, Cano M. Online racism, psychological distress, and alcohol use among racial minority women and men: A multi-group mediation analysis. *Am J Orthopsychiatry* 2021;91(4):524-530. [doi: [10.1037/ort0000553](https://doi.org/10.1037/ort0000553)] [Medline: [34338543](https://pubmed.ncbi.nlm.nih.gov/34338543/)]
36. Daniels J. Race and racism in Internet Studies: A review and critique. *New Media & Society* 2012 Dec 10;15(5):695-719. [doi: [10.1177/1461444812462849](https://doi.org/10.1177/1461444812462849)]
37. Tynes BM, Giang MT, Williams DR, Thompson GN. Online racial discrimination and psychological adjustment among adolescents. *J Adolesc Health* 2008 Dec;43(6):565-569. [doi: [10.1016/j.jadohealth.2008.08.021](https://doi.org/10.1016/j.jadohealth.2008.08.021)] [Medline: [19027644](https://pubmed.ncbi.nlm.nih.gov/19027644/)]

38. Tynes BM, English D, Del Toro J, Smith NA, Lozada FT, Williams DR. Trajectories of Online Racial Discrimination and Psychological Functioning Among African American and Latino Adolescents. *Child Dev* 2020 Sep 14;91(5):1577-1593 [FREE Full text] [doi: [10.1111/cdev.13350](https://doi.org/10.1111/cdev.13350)] [Medline: [31943164](https://pubmed.ncbi.nlm.nih.gov/31943164/)]
39. Keum BT, Miller MJ. Racism in digital era: Development and initial validation of the Perceived Online Racism Scale (PORS v1.0). *J Couns Psychol* 2017 Apr;64(3):310-324. [doi: [10.1037/cou0000205](https://doi.org/10.1037/cou0000205)] [Medline: [28383967](https://pubmed.ncbi.nlm.nih.gov/28383967/)]
40. TaeHyuk Keum B, Hearn M. Online Gaming and Racism: Impact on Psychological Distress Among Black, Asian, and Latinx Emerging Adults. *Games and Culture* 2021 Aug 16;17(3):445-460. [doi: [10.1177/15554120211039082](https://doi.org/10.1177/15554120211039082)]
41. Sue S, Okazaki S. Asian-American educational achievements: A phenomenon in search of an explanation. *American Psychologist* 1990;45(8):913-920. [doi: [10.1037/0003-066x.45.8.913](https://doi.org/10.1037/0003-066x.45.8.913)]
42. Qin DB, Way N, Mukherjee P. The Other Side of the Model Minority Story. *Youth & Society* 2007 Aug 09;39(4):480-506. [doi: [10.1177/0044118x08314233](https://doi.org/10.1177/0044118x08314233)]
43. Haokip T. From 'Chinky' to 'Coronavirus': racism against Northeast Indians during the Covid-19 pandemic. *Asian Ethnicity* 2020 May 18;22(2):353-373. [doi: [10.1080/14631369.2020.1763161](https://doi.org/10.1080/14631369.2020.1763161)]
44. Li Y, Nicholson HL. When "model minorities" become "yellow peril"-Othering and the racialization of Asian Americans in the COVID-19 pandemic. *Sociol Compass* 2021 Feb 16;15(2):e12849 [FREE Full text] [doi: [10.1111/soc4.12849](https://doi.org/10.1111/soc4.12849)] [Medline: [33786062](https://pubmed.ncbi.nlm.nih.gov/33786062/)]
45. Covid-19 Fueling Anti-Asian Racism and Xenophobia Worldwide. Human Rights Watch. 2920. URL: <https://www.hrw.org/news/2020/05/12/covid-19-fueling-anti-asian-racism-and-xenophobia-worldwide#> [accessed 2022-09-08]
46. Cheng T, Conca-Cheng A. The Pandemics of Racism and COVID-19: Danger and Opportunity. *Pediatrics* 2020 Nov;146(5):e2020024836. [doi: [10.1542/peds.2020-024836](https://doi.org/10.1542/peds.2020-024836)] [Medline: [32873718](https://pubmed.ncbi.nlm.nih.gov/32873718/)]
47. Gover AR, Harper SB, Langton L. Anti-Asian Hate Crime During the COVID-19 Pandemic: Exploring the Reproduction of Inequality. *Am J Crim Justice* 2020 Jul 07;45(4):647-667 [FREE Full text] [doi: [10.1007/s12103-020-09545-1](https://doi.org/10.1007/s12103-020-09545-1)] [Medline: [32837171](https://pubmed.ncbi.nlm.nih.gov/32837171/)]
48. Le TK, Cha L, Han H, Tseng W. Anti-Asian Xenophobia and Asian American COVID-19 Disparities. *Am J Public Health* 2020 Sep;110(9):1371-1373. [doi: [10.2105/ajph.2020.305846](https://doi.org/10.2105/ajph.2020.305846)]
49. Misra S, Le PD, Goldmann E, Yang LH. Psychological impact of anti-Asian stigma due to the COVID-19 pandemic: A call for research, practice, and policy responses. *Psychol Trauma* 2020 Jul;12(5):461-464 [FREE Full text] [doi: [10.1037/tra0000821](https://doi.org/10.1037/tra0000821)] [Medline: [32525390](https://pubmed.ncbi.nlm.nih.gov/32525390/)]
50. Tessler H, Choi M, Kao G. The Anxiety of Being Asian American: Hate Crimes and Negative Biases During the COVID-19 Pandemic. *Am J Crim Justice* 2020;45(4):636-646 [FREE Full text] [doi: [10.1007/s12103-020-09541-5](https://doi.org/10.1007/s12103-020-09541-5)] [Medline: [32837158](https://pubmed.ncbi.nlm.nih.gov/32837158/)]
51. National Report (through March 2021). Stop AAPI Hate. 2021 May 6. URL: <https://stopaapihate.org/national-report-through-march-2021/> [accessed 2022-09-08]
52. Nguyen TT, Criss S, Dwivedi P, Huang D, Keralis J, Hsu E, et al. Exploring U.S. Shifts in Anti-Asian Sentiment with the Emergence of COVID-19. *Int J Environ Res Public Health* 2020 Sep 25;17(19):7032 [FREE Full text] [doi: [10.3390/ijerph17197032](https://doi.org/10.3390/ijerph17197032)] [Medline: [32993005](https://pubmed.ncbi.nlm.nih.gov/32993005/)]
53. National Report (through June 2021). Stop AAPI Hate. 2021 Aug 12. URL: <https://stopaapihate.org/stop-aapi-hate-national-report-2/> [accessed 2022-09-08]
54. Noh S, Beiser M, Kaspar V, Hou F, Rummens J. Perceived Racial Discrimination, Depression, and Coping: A Study of Southeast Asian Refugees in Canada. *Journal of Health and Social Behavior* 1999 Sep;40(3):193. [doi: [10.2307/2676348](https://doi.org/10.2307/2676348)]
55. Carter RT. Racism and Psychological and Emotional Injury. *The Counseling Psychologist* 2016 Jun 30;35(1):13-105. [doi: [10.1177/0011000006292033](https://doi.org/10.1177/0011000006292033)]
56. Gee GC, Spencer MS, Chen J, Takeuchi D. A Nationwide Study of Discrimination and Chronic Health Conditions Among Asian Americans. *Am J Public Health* 2007 Jul;97(7):1275-1282. [doi: [10.2105/ajph.2006.091827](https://doi.org/10.2105/ajph.2006.091827)]
57. Harrell SP. A multidimensional conceptualization of racism-related stress: implications for the well-being of people of color. *Am J Orthopsychiatry* 2000 Jan;70(1):42-57. [doi: [10.1037/h0087722](https://doi.org/10.1037/h0087722)] [Medline: [10702849](https://pubmed.ncbi.nlm.nih.gov/10702849/)]
58. Williams DR, Williams-Morris R. Racism and mental health: the African American experience. *Ethn Health* 2000 Jul;5(3-4):243-268. [doi: [10.1080/713667453](https://doi.org/10.1080/713667453)] [Medline: [11105267](https://pubmed.ncbi.nlm.nih.gov/11105267/)]
59. Lee DL, Ahn S. Racial Discrimination and Asian Mental Health: A Meta-Analysis. *The Counseling Psychologist* 2011 Mar 18;39(3):463-489. [doi: [10.1177/0011000010381791](https://doi.org/10.1177/0011000010381791)]
60. Tynes BM, Rose CA, Markoe SL. Extending campus life to the Internet: Social media, discrimination, and perceptions of racial climate. *Journal of Diversity in Higher Education* 2013 Jun;6(2):102-114. [doi: [10.1037/a0033267](https://doi.org/10.1037/a0033267)]
61. Beiser M, Hou F. Mental Health Effects of Premigration Trauma and Postmigration Discrimination on Refugee Youth in Canada. *J Nerv Ment Dis* 2016;204(6):464-470. [doi: [10.1097/nmd.0000000000000516](https://doi.org/10.1097/nmd.0000000000000516)]
62. Carter RT, Lau MY, Johnson V, Kirkinis K. Racial Discrimination and Health Outcomes Among Racial/Ethnic Minorities: A Meta-Analytic Review. *Journal of Multicultural Counseling and Development* 2017 Oct 05;45(4):232-259. [doi: [10.1002/jmcd.12076](https://doi.org/10.1002/jmcd.12076)]
63. English D, Lambert SF, Tynes BM, Bowleg L, Zea MC, Howard LC. Daily multidimensional racial discrimination among Black U.S. American adolescents. *J Appl Dev Psychol* 2020 Jan;66:101068 [FREE Full text] [doi: [10.1016/j.appdev.2019.101068](https://doi.org/10.1016/j.appdev.2019.101068)] [Medline: [33994610](https://pubmed.ncbi.nlm.nih.gov/33994610/)]

64. Serafini K, Coyer C, Brown Speights J, Donovan D, Guh J, Washington J, et al. Racism as Experienced by Physicians of Color in the Health Care Setting. *Fam Med* 2020 Apr;52(4):282-287 [FREE Full text] [doi: [10.22454/FamMed.2020.384384](https://doi.org/10.22454/FamMed.2020.384384)] [Medline: [32267524](https://pubmed.ncbi.nlm.nih.gov/32267524/)]
65. Valkenburg PM, Peter J. The Differential Susceptibility to Media Effects Model. *J Commun* 2013 Mar 07;63(2):221-243. [doi: [10.1111/jcom.12024](https://doi.org/10.1111/jcom.12024)]
66. Broman CL. Race Identity. In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Amsterdam, The Netherlands: Elsevier; 2015:833-836.
67. Lee RM. Resilience Against Discrimination: Ethnic Identity and Other-Group Orientation as Protective Factors for Korean Americans. *Journal of Counseling Psychology* 2005 Jan;52(1):36-44. [doi: [10.1037/0022-0167.52.1.36](https://doi.org/10.1037/0022-0167.52.1.36)]
68. Phinney JS, Ong AD. Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology* 2007 Jul;54(3):271-281. [doi: [10.1037/0022-0167.54.3.271](https://doi.org/10.1037/0022-0167.54.3.271)]
69. Cokley KO. Racial(ized) Identity, Ethnic Identity, and Afrocentric Values: Conceptual and Methodological Challenges in Understanding African American Identity. *Journal of Counseling Psychology* 2005 Oct;52(4):517-526. [doi: [10.1037/0022-0167.52.4.517](https://doi.org/10.1037/0022-0167.52.4.517)]
70. CROSS BE. New Racism, Reformed Teacher Education, and the Same Ole' Oppression. *Educational Studies* 2005 Dec;38(3):263-274. [doi: [10.1207/s15326993es3803\\_6](https://doi.org/10.1207/s15326993es3803_6)]
71. Yoo HC, Lee RM. Ethnic Identity and Approach-Type Coping as Moderators of the Racial Discrimination/Well-Being Relation in Asian Americans. *Journal of Counseling Psychology* 2005 Oct;52(4):497-506. [doi: [10.1037/0022-0167.52.4.497](https://doi.org/10.1037/0022-0167.52.4.497)]
72. Iwamoto DK, Liu WM. The impact of racial identity, ethnic identity, asian values and race-related stress on Asian Americans and Asian international college students' psychological well-being. *J Couns Psychol* 2010 Jan;57(1):79-91 [FREE Full text] [doi: [10.1037/a0017393](https://doi.org/10.1037/a0017393)] [Medline: [20396592](https://pubmed.ncbi.nlm.nih.gov/20396592/)]
73. Chae DH, Takeuchi DT, Barbeau EM, Bennett GG, Lindsey J, Krieger N. Unfair Treatment, Racial/Ethnic Discrimination, Ethnic Identification, and Smoking Among Asian Americans in the National Latino and Asian American Study. *Am J Public Health* 2008 Mar;98(3):485-492. [doi: [10.2105/ajph.2006.102012](https://doi.org/10.2105/ajph.2006.102012)]
74. Yoo HC, Lee RM. Does ethnic identity buffer or exacerbate the effects of frequent racial discrimination on situational well-being of Asian Americans? *Asian American Journal of Psychology* 2009 Aug;S(1):70-87. [doi: [10.1037/1948-1985.s.1.70](https://doi.org/10.1037/1948-1985.s.1.70)]
75. Lee JP, Lee RM, Hu AW, Kim OM. Ethnic Identity as a Moderator against Discrimination for Transracially and Transnationally Adopted Korean American Adolescents. *Asian Am J Psychol* 2015 Jun;6(2):154-163 [FREE Full text] [doi: [10.1037/a0038360](https://doi.org/10.1037/a0038360)] [Medline: [26273427](https://pubmed.ncbi.nlm.nih.gov/26273427/)]
76. Tynes BM, Umaña-Taylor AJ, Rose CA, Lin J, Anderson CJ. Online racial discrimination and the protective function of ethnic identity and self-esteem for African American adolescents. *Dev Psychol* 2012 Mar;48(2):343-355. [doi: [10.1037/a0027032](https://doi.org/10.1037/a0027032)] [Medline: [22369340](https://pubmed.ncbi.nlm.nih.gov/22369340/)]
77. Woo B, Fan W, Tran TV, Takeuchi DT. The role of racial/ethnic identity in the association between racial discrimination and psychiatric disorders: A buffer or exacerbator? *SSM Popul Health* 2019 Apr;7:100378 [FREE Full text] [doi: [10.1016/j.ssmph.2019.100378](https://doi.org/10.1016/j.ssmph.2019.100378)] [Medline: [30923732](https://pubmed.ncbi.nlm.nih.gov/30923732/)]
78. Yip T, Wang Y, Mootoo C, Mirpuri S. Moderating the association between discrimination and adjustment: A meta-analysis of ethnic/racial identity. *Dev Psychol* 2019 Jun;55(6):1274-1298 [FREE Full text] [doi: [10.1037/dev0000708](https://doi.org/10.1037/dev0000708)] [Medline: [30907605](https://pubmed.ncbi.nlm.nih.gov/30907605/)]
79. Peterson LM, Stock ML, Monroe J, Molloy-Paolillo BK, Lambert SF. Racial exclusion causes acute cortisol release among emerging-adult African Americans: The role of reduced perceived control. *J Soc Psychol* 2020 Sep 02;160(5):658-674 [FREE Full text] [doi: [10.1080/00224545.2020.1729685](https://doi.org/10.1080/00224545.2020.1729685)] [Medline: [32122284](https://pubmed.ncbi.nlm.nih.gov/32122284/)]
80. Priest N, Paradies Y, Stevens M, Bailie R. Exploring relationships between racism, housing and child illness in remote indigenous communities. *J Epidemiol Community Health* 2012 May 30;66(5):440-447. [doi: [10.1136/jech.2010.117366](https://doi.org/10.1136/jech.2010.117366)] [Medline: [21118951](https://pubmed.ncbi.nlm.nih.gov/21118951/)]
81. Gibbons FX, Kingsbury JH, Weng C, Gerrard M, Cutrona C, Wills TA, et al. Effects of perceived racial discrimination on health status and health behavior: a differential mediation hypothesis. *Health Psychol* 2014 Jan;33(1):11-19 [FREE Full text] [doi: [10.1037/a0033857](https://doi.org/10.1037/a0033857)] [Medline: [24417690](https://pubmed.ncbi.nlm.nih.gov/24417690/)]
82. Gibbons F, Stock M. Perceived Racial Discrimination and Health Behavior: Mediation and Moderation. In: Major B, Dovidio JF, Link BG, editors. *The Oxford Handbook of Stigma, Discrimination, and Health*. Oxford, UK: Oxford University Press; 2017:355-378.
83. Simons RL, Chen Y, Stewart EA, Brody GH. Incidents of discrimination and risk for delinquency: A longitudinal test of strain theory with an African American sample. *Justice Quarterly* 2003 Dec;20(4):827-854. [doi: [10.1080/07418820300095711](https://doi.org/10.1080/07418820300095711)]
84. Bowman M. Predictors of Reactive Aggression among African American and European American children: the Role of Perceived discrimination, Racial socialization, and Negative Affect. ProQuest. 2022. URL: <https://www.proquest.com/openview/6a8ed59202f12920879bd50a88765844/1?pq-origsite=gscholar&cbl=18750&diss=y> [accessed 2022-09-09]

85. Zhao N, Zhou G. Social Media Use and Mental Health during the COVID-19 Pandemic: Moderator Role of Disaster Stressor and Mediator Role of Negative Affect. *Appl Psychol Health Well Being* 2020 Dec 17;12(4):1019-1038 [FREE Full text] [doi: [10.1111/aphw.12226](https://doi.org/10.1111/aphw.12226)] [Medline: [32945123](https://pubmed.ncbi.nlm.nih.gov/32945123/)]
86. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 1988;54(6):1063-1070. [doi: [10.1037/0022-3514.54.6.1063](https://doi.org/10.1037/0022-3514.54.6.1063)]
87. 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](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]
88. 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](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
89. Bride BE, Robinson MM, Yegidis B, Figley CR. Development and Validation of the Secondary Traumatic Stress Scale. *Research on Social Work Practice* 2016 Aug 19;14(1):27-35. [doi: [10.1177/1049731503254106](https://doi.org/10.1177/1049731503254106)]
90. Yang C, Tsai J, Pan S. Discrimination and Well-Being Among Asians/Asian Americans During COVID-19: The Role of Social Media. *Cyberpsychol Behav Soc Netw* 2020 Dec 01;23(12):865-870. [doi: [10.1089/cyber.2020.0394](https://doi.org/10.1089/cyber.2020.0394)] [Medline: [32762541](https://pubmed.ncbi.nlm.nih.gov/32762541/)]
91. Tynes B, Rose C, Williams D. The Development and Validation of the Online Victimization Scale for Adolescents. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 2010;4(2):Article 2 [FREE Full text]
92. Oyserman D, Sakamoto I. Being Asian American. *The Journal of Applied Behavioral Science* 2016 Jul 26;33(4):435-453. [doi: [10.1177/0021886397334002](https://doi.org/10.1177/0021886397334002)]
93. Chao G, Emlaw J, Chiu A, Yang J, Thumma J, Brackett A, et al. Asian American Pacific Islander Representation in Outcomes Research: NSQIP Scoping Review. *J Am Coll Surg* 2021 May;232(5):682-689.e5. [doi: [10.1016/j.jamcollsurg.2021.01.015](https://doi.org/10.1016/j.jamcollsurg.2021.01.015)] [Medline: [33705984](https://pubmed.ncbi.nlm.nih.gov/33705984/)]
94. Hayes A. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (2nd Edition). New York, NY: Guilford Publications; 2017.
95. Hall BJ, Xiong YX, Yip PSY, Lao CK, Shi W, Sou EKL, et al. The association between disaster exposure and media use on post-traumatic stress disorder following Typhoon Hato in Macao, China. *Eur J Psychotraumatol* 2019 Jan 14;10(1):1558709 [FREE Full text] [doi: [10.1080/20008198.2018.1558709](https://doi.org/10.1080/20008198.2018.1558709)] [Medline: [30693078](https://pubmed.ncbi.nlm.nih.gov/30693078/)]
96. Zeballos Rivas DR, Lopez Jaldin ML, Nina Canaviri B, Portugal Escalante LF, Alanes Fernández AMC, Aguilar Ticona JP. Social media exposure, risk perception, preventive behaviors and attitudes during the COVID-19 epidemic in La Paz, Bolivia: A cross sectional study. *PLoS One* 2021 Jan 22;16(1):e0245859 [FREE Full text] [doi: [10.1371/journal.pone.0245859](https://doi.org/10.1371/journal.pone.0245859)] [Medline: [33481945](https://pubmed.ncbi.nlm.nih.gov/33481945/)]
97. Hossain MT, Ahammed B, Chanda SK, Jahan N, Ela MZ, Islam MN. Social and electronic media exposure and generalized anxiety disorder among people during COVID-19 outbreak in Bangladesh: A preliminary observation. *PLoS One* 2020 Sep 11;15(9):e0238974 [FREE Full text] [doi: [10.1371/journal.pone.0238974](https://doi.org/10.1371/journal.pone.0238974)] [Medline: [32916691](https://pubmed.ncbi.nlm.nih.gov/32916691/)]
98. Patel M, Kute V, Agarwal S. "Infodemic" of COVID 19: More pandemic than the virus. *Indian J Nephrol* 2020;30(3):188. [doi: [10.4103/ijn.ijn\\_216\\_20](https://doi.org/10.4103/ijn.ijn_216_20)]
99. Zack E, Kennedy J, Long J. Can Nonprobability Samples Be Used for Social Science Research? a Cautionary Tale. *Survey Research Methods* 2019;13(2):215-227 [FREE Full text] [doi: [10.18148/srm/2019.v13i2.7262](https://doi.org/10.18148/srm/2019.v13i2.7262)]
100. Hu Z, Yang Z, Li Q, Zhang A. The COVID-19 Infodemic: Infodemiology Study Analyzing Stigmatizing Search Terms. *J Med Internet Res* 2020 Nov 16;22(11):e22639 [FREE Full text] [doi: [10.2196/22639](https://doi.org/10.2196/22639)] [Medline: [33156807](https://pubmed.ncbi.nlm.nih.gov/33156807/)]
101. Rathore F, Farooq F. Information Overload and Infodemic in the COVID-19 Pandemic. *J Pak Med Assoc* 2020;1. [doi: [10.5455/jpma.38](https://doi.org/10.5455/jpma.38)]
102. Pandya A, Lodha P. Social Connectedness, Excessive Screen Time During COVID-19 and Mental Health: A Review of Current Evidence. *Front. Hum. Dyn* 2021 Jul 22;3:3. [doi: [10.3389/fhumd.2021.684137](https://doi.org/10.3389/fhumd.2021.684137)]
103. Király O, Potenza MN, Stein DJ, King DL, Hodgins DC, Saunders JB, et al. Preventing problematic internet use during the COVID-19 pandemic: Consensus guidance. *Compr Psychiatry* 2020 Jul;100:152180 [FREE Full text] [doi: [10.1016/j.comppsy.2020.152180](https://doi.org/10.1016/j.comppsy.2020.152180)] [Medline: [32422427](https://pubmed.ncbi.nlm.nih.gov/32422427/)]
104. King D, Delfabbro P, Billieux J, Potenza M. Problematic online gaming and the COVID-19 pandemic. *J Behav Addict* 2020 Jun;9(2):184-186 [FREE Full text] [doi: [10.1556/2006.2020.00016](https://doi.org/10.1556/2006.2020.00016)] [Medline: [32352927](https://pubmed.ncbi.nlm.nih.gov/32352927/)]
105. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychol Sci* 2020 Jul 30;31(7):770-780 [FREE Full text] [doi: [10.1177/0956797620939054](https://doi.org/10.1177/0956797620939054)] [Medline: [32603243](https://pubmed.ncbi.nlm.nih.gov/32603243/)]
106. Centers for Disease Control and Prevention (CDC). COVID Data Tracker. CDC. 2020 Mar 28. URL: <https://covid.cdc.gov/covid-data-tracker/#demographicsovertime> [accessed 2022-09-09]
107. Kim EJ, Marrast L, Conigliaro J. COVID-19: Magnifying the Effect of Health Disparities. *J Gen Intern Med* 2020 Aug 11;35(8):2441-2442 [FREE Full text] [doi: [10.1007/s11606-020-05881-4](https://doi.org/10.1007/s11606-020-05881-4)] [Medline: [32394141](https://pubmed.ncbi.nlm.nih.gov/32394141/)]
108. Lopez L, Hart LH, Katz MH. Racial and Ethnic Health Disparities Related to COVID-19. *JAMA* 2021 Feb 23;325(8):719-720. [doi: [10.1001/jama.2020.26443](https://doi.org/10.1001/jama.2020.26443)] [Medline: [33480972](https://pubmed.ncbi.nlm.nih.gov/33480972/)]

109. Gomez A, Grantham-Philips W, Hughes T. “An unbelievable chain of oppression”: America’s history of racism was a preexisting condition for COVID-19. USA Today. 2020. URL: <https://www.usatoday.com/in-depth/news/nation/2020/10/12/coronavirus-deaths-reveal-systemic-racism-united-states/5770952002/> [accessed 2022-09-09]
110. Elbaum A. Black Lives in a Pandemic: Implications of Systemic Injustice for End-of-Life Care. *Hastings Cent Rep* 2020 May 29;50(3):58-60 [FREE Full text] [doi: [10.1002/hast.1135](https://doi.org/10.1002/hast.1135)] [Medline: [32596896](https://pubmed.ncbi.nlm.nih.gov/32596896/)]
111. Yancy CW. COVID-19 and African Americans. *JAMA* 2020 May 19;323(19):1891-1892. [doi: [10.1001/jama.2020.6548](https://doi.org/10.1001/jama.2020.6548)] [Medline: [32293639](https://pubmed.ncbi.nlm.nih.gov/32293639/)]
112. Egede LE, Walker RJ. Structural Racism, Social Risk Factors, and Covid-19 - A Dangerous Convergence for Black Americans. *N Engl J Med* 2020 Sep 17;383(12):e77 [FREE Full text] [doi: [10.1056/NEJMp2023616](https://doi.org/10.1056/NEJMp2023616)] [Medline: [32706952](https://pubmed.ncbi.nlm.nih.gov/32706952/)]
113. Centers for Disease Control and Prevention (CDC). Risk for Covid-19 Infection, Hospitalization, and Death by Race/Ethnicity. CDC. 2020. URL: <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html> [accessed 2022-09-09]
114. Noe-Bustamante L, Krogstad JM, Lopez MH. Centers for Disease Control and Prevention. Pew Research Center. 2021 Jul 15. URL: <https://www.pewresearch.org/race-ethnicity/2021/07/15/for-u-s-latinos-covid-19-has-taken-a-personal-and-financial-toll/> [accessed 2022-09-09]
115. Arrazola J, Masiello MM, Joshi S, Dominguez AE, Poel A, Wilkie CM, et al. COVID-19 Mortality Among American Indian and Alaska Native Persons - 14 States, January-June 2020. *MMWR Morb Mortal Wkly Rep* 2020 Dec 11;69(49):1853-1856 [FREE Full text] [doi: [10.15585/mmwr.mm6949a3](https://doi.org/10.15585/mmwr.mm6949a3)] [Medline: [33301432](https://pubmed.ncbi.nlm.nih.gov/33301432/)]
116. Leggat-Barr K, Uchikoshi F, Goldman N. COVID-19 risk factors and mortality among Native Americans. *DemRes* 2021 Nov 17;45:1185-1218. [doi: [10.4054/demres.2021.45.39](https://doi.org/10.4054/demres.2021.45.39)]
117. Acosta AM, Garg S, Pham H, Whitaker M, Anglin O, O'Halloran A, et al. Racial and Ethnic Disparities in Rates of COVID-19-Associated Hospitalization, Intensive Care Unit Admission, and In-Hospital Death in the United States From March 2020 to February 2021. *JAMA Netw Open* 2021 Oct 01;4(10):e2130479 [FREE Full text] [doi: [10.1001/jamanetworkopen.2021.30479](https://doi.org/10.1001/jamanetworkopen.2021.30479)] [Medline: [34673962](https://pubmed.ncbi.nlm.nih.gov/34673962/)]
118. Bleich SN, Findling MG, Casey LS, Blendon RJ, Benson JM, SteelFisher GK, et al. Discrimination in the United States: Experiences of black Americans. *Health Serv Res* 2019 Dec 29;54 Suppl 2(S2):1399-1408 [FREE Full text] [doi: [10.1111/1475-6773.13220](https://doi.org/10.1111/1475-6773.13220)] [Medline: [31663124](https://pubmed.ncbi.nlm.nih.gov/31663124/)]
119. Library of Congress. A Century of Racial Segregation 1849–1950 - Brown v. Board at fifty: “With an Even hand”. Library of Congress. 1950. URL: <https://www.loc.gov/exhibits/brown/brown-segregation.html> [accessed 2022-09-09]
120. Bernstein K, Ahern J, Tracy M, Boscarino J, Vlahov D, Galea S. Television watching and the risk of incident probable posttraumatic stress disorder: a prospective evaluation. *J Nerv Ment Dis* 2007 Jan;195(1):41-47. [doi: [10.1097/01.nmd.0000244784.36745.a5](https://doi.org/10.1097/01.nmd.0000244784.36745.a5)] [Medline: [17220738](https://pubmed.ncbi.nlm.nih.gov/17220738/)]
121. Laurencin CT, Walker JM. A Pandemic on a Pandemic: Racism and COVID-19 in Blacks. *Cell Syst* 2020 Jul 22;11(1):9-10 [FREE Full text] [doi: [10.1016/j.cels.2020.07.002](https://doi.org/10.1016/j.cels.2020.07.002)] [Medline: [32702320](https://pubmed.ncbi.nlm.nih.gov/32702320/)]
122. Millett GA, Jones AT, Benkeser D, Baral S, Mercer L, Beyrer C, et al. Assessing differential impacts of COVID-19 on black communities. *Ann Epidemiol* 2020 Jul;47:37-44 [FREE Full text] [doi: [10.1016/j.annepidem.2020.05.003](https://doi.org/10.1016/j.annepidem.2020.05.003)] [Medline: [32419766](https://pubmed.ncbi.nlm.nih.gov/32419766/)]
123. Caporuscio J. What Is White fragility, and Why Is It a problem? *www.medicalnewstoday*. 2020 Jun 12. URL: <https://www.medicalnewstoday.com/articles/white-fragility-definition> [accessed 2022-04-08]
124. Diangelo RJ. *White Fragility: Why It’s so Hard for White People to Talk about Racism*. Boston, MA: Beacon Press; 2018.
125. Smith J. Medical care at protests must be defended. *J Public Health (Oxf)* 2021 Jun 07;43(2):e261. [doi: [10.1093/pubmed/fdaa191](https://doi.org/10.1093/pubmed/fdaa191)] [Medline: [33164086](https://pubmed.ncbi.nlm.nih.gov/33164086/)]
126. Thomas SP. Black Mental Health Matters: Addressing Post-COVID Mental Health Needs of Black Americans. *Issues Ment Health Nurs* 2021 Aug 27;42(8):707-708. [doi: [10.1080/01612840.2021.1952017](https://doi.org/10.1080/01612840.2021.1952017)] [Medline: [34314663](https://pubmed.ncbi.nlm.nih.gov/34314663/)]
127. Saltzman LY, Lesen AE, Henry V, Hansel TC, Bordnick PS. COVID-19 Mental Health Disparities. *Health Secur* 2021 Jun 01;19(S1):S5-S13. [doi: [10.1089/hs.2021.0017](https://doi.org/10.1089/hs.2021.0017)] [Medline: [34014118](https://pubmed.ncbi.nlm.nih.gov/34014118/)]
128. Towers S, Afzal S, Bernal G, Bliss N, Brown S, Espinoza B, et al. Mass Media and the Contagion of Fear: The Case of Ebola in America. *PLoS One* 2015 Jun 11;10(6):e0129179 [FREE Full text] [doi: [10.1371/journal.pone.0129179](https://doi.org/10.1371/journal.pone.0129179)] [Medline: [26067433](https://pubmed.ncbi.nlm.nih.gov/26067433/)]
129. Holman EA, Garfin DR, Lubens P, Silver RC. Media Exposure to Collective Trauma, Mental Health, and Functioning: Does It Matter What You See? *Clinical Psychological Science* 2019 Oct 08;8(1):111-124. [doi: [10.1177/2167702619858300](https://doi.org/10.1177/2167702619858300)]
130. Garfin DR, Holman EA, Silver RC. Cumulative exposure to prior collective trauma and acute stress responses to the Boston marathon bombings. *Psychol Sci* 2015 Jun 20;26(6):675-683. [doi: [10.1177/0956797614561043](https://doi.org/10.1177/0956797614561043)] [Medline: [25896419](https://pubmed.ncbi.nlm.nih.gov/25896419/)]
131. Hopwood TL, Schutte NS. Psychological outcomes in reaction to media exposure to disasters and large-scale violence: A meta-analysis. *Psychology of Violence* 2017;7(2):316-327. [doi: [10.1037/vio0000056](https://doi.org/10.1037/vio0000056)]
132. Schlenger WE, Caddell JM, Ebert L, Jordan BK, Rourke KM, Wilson D, et al. Psychological reactions to terrorist attacks: findings from the National Study of Americans’ Reactions to September 11. *JAMA* 2002 Aug 07;288(5):581-588. [doi: [10.1001/jama.288.5.581](https://doi.org/10.1001/jama.288.5.581)] [Medline: [12150669](https://pubmed.ncbi.nlm.nih.gov/12150669/)]

133. Laugharne J, Janca A, Widiger T. Posttraumatic stress disorder and terrorism: 5 years after 9/11. *Current Opinion in Psychiatry* 2007;20(1):36-41. [doi: [10.1097/ycp.0b013e328010dc2c](https://doi.org/10.1097/ycp.0b013e328010dc2c)]
134. Holman EA, Garfin DR, Silver RC. Media's role in broadcasting acute stress following the Boston Marathon bombings. *Proc Natl Acad Sci U S A* 2014 Jan 07;111(1):93-98 [FREE Full text] [doi: [10.1073/pnas.1316265110](https://doi.org/10.1073/pnas.1316265110)] [Medline: [24324161](https://pubmed.ncbi.nlm.nih.gov/24324161/)]
135. Jones NM, Garfin DR, Holman EA, Silver RC. Media Use and Exposure to Graphic Content in the Week Following the Boston Marathon Bombings. *Am J Community Psychol* 2016 Sep 12;58(1-2):47-59. [doi: [10.1002/ajcp.12073](https://doi.org/10.1002/ajcp.12073)] [Medline: [27616665](https://pubmed.ncbi.nlm.nih.gov/27616665/)]
136. Robert M, Stene L, Garfin D, Vandentorren S, Motreff Y, du Roscoat E, et al. Media Exposure and Post-traumatic Stress Symptoms in the Wake of the November 2015 Paris Terrorist Attacks: A Population-Based Study in France. *Front Psychiatry* 2021;12:509457 [FREE Full text] [doi: [10.3389/fpsy.2021.509457](https://doi.org/10.3389/fpsy.2021.509457)] [Medline: [34093248](https://pubmed.ncbi.nlm.nih.gov/34093248/)]
137. Schmitt MT, Branscombe NR, Postmes T, Garcia A. The consequences of perceived discrimination for psychological well-being: a meta-analytic review. *Psychol Bull* 2014 Jul;140(4):921-948. [doi: [10.1037/a0035754](https://doi.org/10.1037/a0035754)] [Medline: [24547896](https://pubmed.ncbi.nlm.nih.gov/24547896/)]
138. Mayrl D, Saperstein A. When white people report racial discrimination: The role of region, religion, and politics. *Soc Sci Res* 2013 May;42(3):742-754. [doi: [10.1016/j.ssresearch.2012.12.007](https://doi.org/10.1016/j.ssresearch.2012.12.007)] [Medline: [23521992](https://pubmed.ncbi.nlm.nih.gov/23521992/)]
139. Craig MA, Richeson JA. Information about the US racial demographic shift triggers concerns about anti-White discrimination among the prospective White "minority". *PLoS One* 2017 Sep 27;12(9):e0185389 [FREE Full text] [doi: [10.1371/journal.pone.0185389](https://doi.org/10.1371/journal.pone.0185389)] [Medline: [28953971](https://pubmed.ncbi.nlm.nih.gov/28953971/)]
140. Clark R, Anderson NB, Clark VR, Williams DR. Racism as a stressor for African Americans: A biopsychosocial model. *American Psychologist* 1999;54(10):805-816. [doi: [10.1037/0003-066x.54.10.805](https://doi.org/10.1037/0003-066x.54.10.805)]
141. Sellers RM, Caldwell CH, Schmeelk-Cone KH, Zimmerman MA. Racial Identity, Racial Discrimination, Perceived Stress, and Psychological Distress among African American Young Adults. *Journal of Health and Social Behavior* 2003 Sep;44(3):302. [doi: [10.2307/1519781](https://doi.org/10.2307/1519781)]
142. Chandler J, Rosenzweig C, Moss AJ, Robinson J, Litman L. Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behav Res Methods* 2019 Oct 11;51(5):2022-2038 [FREE Full text] [doi: [10.3758/s13428-019-01273-7](https://doi.org/10.3758/s13428-019-01273-7)] [Medline: [31512174](https://pubmed.ncbi.nlm.nih.gov/31512174/)]
143. Walters K, Christakis DA, Wright DR. Are Mechanical Turk worker samples representative of health status and health behaviors in the U.S.? *PLoS ONE* 2018 Jun 7;13(6):e0198835. [doi: [10.1371/journal.pone.0198835](https://doi.org/10.1371/journal.pone.0198835)]

## Abbreviations

- AAPI:** Asian, Pacific Islander, and Mixed Asian
- DSMM:** Differential Susceptibility to Media Effects Model
- IE:** indirect effect
- MTurk:** Mechanical Turk
- PSMU:** problematic social media usage
- STS:** secondary traumatic stress

*Edited by A Mavragani; submitted 07.04.22; peer-reviewed by M Pritchard, A Mukherjea; comments to author 21.06.22; revised version received 13.07.22; accepted 25.08.22; published 19.09.22.*

*Please cite as:*

Layug A, Krishnamurthy S, McKenzie R, Feng B

*The Impacts of Social Media Use and Online Racial Discrimination on Asian American Mental Health: Cross-sectional Survey in the United States During COVID-19*

*JMIR Form Res* 2022;6(9):e38589

URL: <https://formative.jmir.org/2022/9/e38589>

doi: [10.2196/38589](https://doi.org/10.2196/38589)

PMID: [36121698](https://pubmed.ncbi.nlm.nih.gov/36121698/)

©Alyan Layug, Samiksha Krishnamurthy, Rachel McKenzie, Bo Feng. Originally published in *JMIR Formative Research* (<https://formative.jmir.org>), 19.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Formative Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.



Original Paper

# Digital Storytelling Methods to Empower Young Black Adults in COVID-19 Vaccination Decision-Making: Feasibility Study and Demonstration

Allysha Maragh-Bass<sup>1,2</sup>, MPH, PhD; Maria Leonora Comello<sup>3</sup>, MSc, PhD; Elizabeth Ellen Tolley<sup>1,4</sup>, MA, PhD; Darrell Stevens Jr<sup>5</sup>; Jade Wilson<sup>3</sup>; Christina Toval<sup>6</sup>, MPH; Henna Budhwani<sup>7</sup>, MPH, PhD; Lisa Hightow-Weidman<sup>4,6</sup>, MD, MPH

<sup>1</sup>Behavioral, Epidemiological, Clinical Sciences Division, FHI 360, Durham, NC, United States

<sup>2</sup>Duke Global Health Institute, Duke University, Durham, NC, United States

<sup>3</sup>Hussman School of Journalism and Media, University of North Carolina, Chapel Hill, NC, United States

<sup>4</sup>Gillings School of Public Health, University of North Carolina, Chapel Hill, NC, United States

<sup>5</sup>DesignLab, FHI 360, Durham, NC, United States

<sup>6</sup>Institute for Global Health and Infectious Diseases, University of North Carolina, Chapel Hill, NC, United States

<sup>7</sup>School of Public Health, University of Alabama at Birmingham, Birmingham, AL, United States

**Corresponding Author:**

Allysha Maragh-Bass, MPH, PhD

Behavioral, Epidemiological, Clinical Sciences Division, FHI 360

359 Blackwell Street #200

Durham, NC, 27701

United States

Phone: 1 919 544 7040 ext 11966

Fax: 1 954 442 9233

Email: [AMaraghBass@fhi360.org](mailto:AMaraghBass@fhi360.org)

## Abstract

**Background:** Despite high rates of novel COVID-19, acceptance of COVID-19 vaccination is low among Black adults. In response, we developed a digital health intervention (Tough Talks-COVID) that includes digital stories created in a workshop we held with young Black adults.

**Objective:** Our formative research using digital storytelling workshops asked 3 research questions: (1) What issues did participants have in conceptualizing their stories, and what themes emerged from the stories they created? (2) What issues did participants have related to production techniques, and which techniques were utilized in stories? and (3) Overall, how did participants evaluate their workshop experience?

**Methods:** Participants were workshop-eligible if they were vaccine-accepting based on a baseline survey fielded in late 2021. Final participants (N=11) completed a consent process, all 3 workshops, and a media release form for their digital story. The first 2 workshops provided background information and hands-on digital storytelling skills from pre- to postproduction. The third workshop served as a screening and feedback session for participants' final videos. Qualitative and quantitative feedback elements were incorporated into all 3 sessions.

**Results:** Digital stories addressed one or more of 4 broad themes: (1) COVID-19 vulnerability, (2) community connections, (3) addressing vaccine hesitancy, and (4) countering vaccine misinformation. Participants incorporated an array of technical approaches, including unique creative elements such as cartoon images and instant messaging tools to convey social interactions around COVID-19 decision-making. Most (9/11, 82%) strongly agreed the digital storytelling workshops were delivered as expected; 10 of 11 agreed (n=5) or strongly agreed (n=5) that they had some ideas about what story to tell by the end of the first workshop, and most (8/11, 73%) strongly agreed they had narrowed down their ideas by workshop two. Of the participants, 9 felt they would very likely (n=6) or likely (n=3) use digital storytelling techniques for personal use in the future, and even more were very likely (n=7) to use the techniques for professional use.

**Conclusions:** Our study is one of the first to incorporate digital storytelling as a central component to a digital health intervention and the only one to do so with exclusive focus on young Black adults. Our emphasis on digital storytelling was shown to be

highly acceptable. Similar approaches, including careful consideration of the ethical challenges of community-based participatory approaches, are applicable to other populations experiencing both COVID-19 inequities and marginalization, such as other age demographics and people of color.

(*JMIR Form Res* 2022;6(9):e38070) doi:[10.2196/38070](https://doi.org/10.2196/38070)

## KEYWORDS

young Black adults; COVID-19; vaccine hesitancy; digital storytelling; community-based participatory research; digital health intervention

## Introduction

### Background

The inequitable impact of the novel COVID-19 pandemic on Black, Indigenous, and other People of Color (BIPOC) communities has manifested itself in myriad ways. Beyond disparities in morbidity and mortality [1,2], the effects have included stigma, discrimination, and distress [3], particularly among those with multiple minority identities [4]. Inequities persist due, in part, to factors like frontline worker positions, higher rates of underlying comorbidities, and limited access to and quality of care [5]. All these challenges predated COVID-19, and in the era of telehealth, the digital divide has exacerbated them [6].

Despite greater likelihood of COVID-19 infection, Black communities have been slower to accept vaccination, due to longstanding historical misrepresentation, exclusion, discrimination, and exploitation in clinical research [7-9]. Furthermore, messages promoting the vaccine—through social and mainstream media outlets—reflect the voices of communities other than their own. Misinformation about COVID-19 vaccination, therefore, has spread widely, and lower rates of acceptance of COVID-19 vaccination among young adults are being seen [10]. Young Black adults, as frequent users of social media [11], have potential to play a crucial role in slowing the spread of misinformation and changing the narrative about the virus. Although they may be exposed to more potential sources of (mis)information, they also have potential to create content that is credible, engaging, and expressive of cultural pride [12].

To address these disparities, our team has developed a digital health intervention (Tough Talks: COVID) with a digital storytelling component featuring short videos created by and for young Black adults. Grounded in real-life experiences, these videos showcase how people have made decisions to get the vaccine and its impact on their lives. We describe digital storytelling workshops in which participants learned skills to create these videos, and we summarize the themes and features of videos, participants' feedback, and personal takeaways. Therefore, this research is formative and documents the process by which digital storytelling approaches were used to empower young Black adults in COVID-19 vaccine decision-making. This work bridges disciplines of digital storytelling and community-based participatory research (CBPR); to our knowledge, it is the only work to do so in the context of vaccine uptake among young Black adults.

### Digital Storytelling and Community-Based Participatory Approaches

We define a story as a representation of thoughts or events with an explicit or implicit message about a topic, in line with other work on health narratives [13]. *Digital* stories often combine multimedia components (eg, video, audio) and are typically shared with wider networks online; the approach is well-suited to the experiences of communities experiencing marginalization due to factors like medical racism [14]. Although digital storytelling is relatively new, use of narratives has a long history in health behavior change [15]. Stories are especially effective when change requires overcoming resistance, processing complex information, and dealing with existential issues [13]—challenges that are all present in the context of preventing the spread of COVID-19 in Black communities. Stories may suspend habitual ways of thinking by eliciting emotion, enjoyment, or identification with characters and, in so doing, generate openness to new perspectives [16]. In the case of digital stories, the social validation of stories that are spread through social media may make them even more compelling.

Our project combines digital storytelling with the well-established approach of CBPR for addressing health inequities [17,18]. A hallmark of CBPR is involvement of stakeholders as partners in the research at all stages of the process, and in response, we engaged young adults as youth advisors to the project and as the creators of digital stories in our intervention itself. CBPR and digital storytelling are complementary approaches because both center the voices of communities, particularly those who experience marginalization. It is an ethical imperative to uplift the voices of these communities, and digital storytelling in the context of CBPR is an innovative approach in line with this. Further, the asset-building emphasis of CBPR [19] is congruent with our goal in digital storytelling workshops of providing skills training and support for digital storytelling. Despite these strengths, inherent to CBPR are several ethical considerations, including how to achieve a true community-driven agenda, how to navigate the tension of insiders versus outsiders, recognizing the limits of participation, and deciding on shared ownership and use of findings [20]. We are mindful of these issues and communicate regularly with our partners on them; in the discussion, we return to these considerations in light of our digital storytelling findings.

### Research Questions

With the goal of empowering young Black adults with the skills to create digital stories about their experiences with COVID-19

and vaccination, we conducted a series of workshops with 2 cohorts. We asked the following research questions (RQs):

- RQ1: What issues did participants have as they conceptualized their stories, and what themes emerged from the stories they created?
- RQ2: What issues did participants have related to production techniques, and which techniques were utilized in stories?
- RQ3: Overall, how did participants evaluate their workshop experience?

## Methods

### Overview

Digital storytelling workshops were conducted as part of a larger study (Tough Talks: COVID) aimed at testing a digital health intervention to empower young Black people in the southern United States to make autonomous decisions about COVID-19 vaccine receipt with culturally tailored and appropriate resources. The entire study (including digital storytelling activities) is guided by a youth advisory board comprised entirely of young Black adults in the 3 regions of the South where the study is being conducted (Alabama [AL], Georgia [GA], and North Carolina [NC]) and an expert advisory board comprised predominantly of Black experts in bioethics, intersectional stigma, health communication, faith-based initiatives, and CBPR. The aims of the larger study are to (1) conduct formative research to elicit the behavioral, cognitive, and environmental determinants influencing COVID-19 vaccine hesitancy among young Black people; (2) utilize implementation science to iteratively develop and refine the Tough Talks-COVID digital intervention with guidance from advisors; and (3) conduct a hybrid type 1 effectiveness implementation trial with young Black people from AL, GA, and NC to assess acceptability, feasibility, and effectiveness of the Tough Talks-COVID intervention on increasing COVID-19 vaccine uptake.

This research addresses the formative research proposed in Aim 1, which includes a series of digital storytelling workshops with young Black adults to create their own digital stories with potential to be included in the Tough Talks-COVID digital health intervention.

### Participants

Digital storytelling workshop participants were recruited through an online survey fielded in the larger study (N=150) from August 2021 to November 2021 and described elsewhere [21]. Respondents all self-identified as Black in survey demographic questions in the larger study from which workshop participants were identified. Median age was 23 (IQR 20-26) years. Over 80% (122/150, 81.3%) were vaccinated (1 dose of a 2-dose vaccine or fully vaccinated). Compared with cisgender women, cisgender men represented 27.3% (41/150) of the overall sample but roughly 40% (11/28, 39%) of unvaccinated individuals. Most participants had some college education or greater (125/150, 83.3%), and among essential workers (N=49), the proportion was roughly 30% (8/49, 16%). Participants were eligible for workshops if they reported willingness to or previous

completion of a COVID-19 vaccine series in the survey. Each survey respondent who was eligible and expressed interest was scheduled for a meeting with a study team coordinator to be consented and address any questions about participation. Our final workshop participants (N=11) were those who completed the survey and all workshop activities. Participants were divided into 2 cohorts to (1) offer 2 sets of dates for which participants could gauge their availability and (2) ensure a small enough group size that all participants were given ample time to work in small groups with our study team when developing their digital stories. With our final group, cohort 1 had 6 participants, and cohort 2 had 5 participants.

### Digital Storytelling Workshops

All workshops took place on a Health Insurance Portability and Accountability Act (HIPAA)-compliant Zoom platform and were scheduled on weekday evenings accommodating work and school schedules. Each workshop lasted between 120 minutes and 150 minutes and were held 2 weeks apart from October 2021 to November 2021. Workshops were led by a group of 5 facilitators, 4 of whom identify as people of color, and 3 of whom identify as young Black adults.

The goal of the series was to help participants develop a digital story (1 minute to 3 minutes) for use in the Tough Talks-COVID intervention. The series was designed to be comprehensive yet approachable to participants with varying levels of experience with digital storytelling. The first 2 workshops provided background information and hands-on digital storytelling skills from pre- to postproduction. The third workshop was a screening and feedback session for participants' final videos, as well as an opportunity to discuss how participants could highlight their digital storytelling workshop experience on their resume and professional lives. Additionally, the third workshop combined participants in both cohorts so that they could serve as audiences for each other's videos. Most but not all participants in our 2 digital storytelling cohorts chose to screen their videos and discuss their experiences during a final "showcase" event attended by community partners and the wider research team.

Because of the virtual format of the sessions, it was important to keep participants engaged. One strategy was the use of ice breakers, such as asking a participant to name their city of birth and favorite food and to "popcorn" it to another participant to do the same until everyone had shared. Another ice breaker involved creating a collaborative story, in which a workshop facilitator began a story with a one-sentence opening and passed it to another person who would then add another sentence, until all facilitators and participants had contributed. Another engagement strategy was the use of emotion cards to gauge participants' comfort level with different activities (see Figure 1). Cards featured 6 licensed stock images of young adults of color and were numbered from 1 to 6, without additional labels but with clearly portrayed emotions (eg, shyness, excitement). In each case, we invited participants to say or type the number corresponding to the emotion and then, in their own words, describe what they would call the emotion. All sessions ended with a recap of key information and any reminders for the next session.

**Figure 1.** Emotion card.

## Sources of Feedback

### *Qualitative (Brainstorming Discussions and Breakout Groups)*

In addition to workshop engagement approaches, we incorporated qualitative feedback and evaluation measures into each workshop. During session one, we utilized Google Jamboards to facilitate discussion around 2 key questions: (1) What COVID-19 vaccine stories need to be told? and (2) What could these stories look and sound like? During session two, the same Jamboard questions were utilized but with more focus on the themes and technical aspects of the “practice” videos they were asked to create. Facilitators with digital storytelling expertise led breakouts to address any specific editing or technical questions. During session three, we utilized Jamboard to facilitate discussion related to 3 questions: (1) How do you feel about the workshops, now that you have completed them; (2) How would you explain the experience to others who may not be familiar with digital storytelling; and (3) How (if at all) could you use these skills moving forward? Team members recorded responses in real time to all questions in the form of “sticky notes” and actively confirmed with participants while writing them that the notes were accurately representing their feedback. Thematic analyses were conducted by team member EET and reviewed and discussed with team members AMB, MLC, and DS to ensure they were grounded in feedback from participants.

### *Quantitative (Feedback Forms)*

At the completion of session three, participants completed a 9-question feedback form about their experience. Each question utilized a 4-point Likert scale and included items such as “The digital storytelling workshops were like what was explained to me during the consent process,” “How likely are you to use digital storytelling in the future for your professional use?” and “After completing the digital storytelling workshops, I feel this was worth my time.” Descriptive statistics were conducted with the total number of participants who completed all 3 workshops (N=11). Between sessions one and two, 3 participants from the original 14 discontinued participation due to work schedules or inability to complete a digital story in time.

## Ethical Considerations

All workshop participants completed (1) a survey, (2) an informed consent discussion with our study coordinator prior to consenting to the workshops with opportunities to ask questions, (3) all 3 workshop sessions, and (4) a media release form consenting to the use of their digital story in the Tough Talks-COVID digital health intervention in addition to a consent form for participation in the workshops. For all sessions, participants were asked to keep cameras on, if preferred, and to share names and pronouns; participants were also able to use preferred names or pseudonyms in all sessions. Participants received US \$50 for each session in which they participated and an additional US \$75 for successful video creation. All digital storytelling workshop activities were fully approved by the University of North Carolina at Chapel Hill Institutional Review Board (IRB# IGHID 12112).

## Results

Across the 2 cohorts of participants in the workshop series, 11 unique videos were created, ranging in content and technical approach and from 1 minute to 3 minutes in length. In the following sections, we detail the workshop sessions and final videos, organized by our 3 research questions on the types of stories told (RQ1), the video and digital techniques used (RQ2), and participants’ overall assessment of the process (RQ3.)

### **RQ 1: Issues Related to Storytelling and Types of Stories Told**

Although some participants were initially anxious about determining what story “should be told,” the digital storytelling process helped most to develop their own unique story about COVID-19 and to answer the question “Why get the vaccine?” Summarized in the following sections and in [Figures 2 and 3](#) are sample “sticky notes” from participant brainstorming in sessions one and two prior to the final digital stories’ creation.

Our summary, therefore, also describes the content ultimately included in videos, which centered around one or more of 4 broad themes, which we have labeled (1) COVID-19 vulnerability, (2) community connections, (3) addressing vaccine hesitancy, and (4) countering vaccine misinformation.

Figure 2. Digital story content brainstorming and development in workshop cohort #1. NCCU: North Carolina Central University.



Figure 3. Digital story content brainstorming and development with workshop cohort #2.



**COVID-19 Vulnerabilities**

As participants worked through initial story ideas in digital storytelling workshop sessions one and two, fear of getting COVID-19 was a common theme. Participants suggested that fear (of getting COVID, of losing a loved one to the disease, or of concerns that chronic underlying conditions could make someone more vulnerable to COVID-19 death or long-term health effects) was a strong motivating factor for vaccine uptake. Within this theme, a range of stories emerged, including a

first-person experience of having gotten COVID-19 and the emotional and physical effects experienced; a story about a friend who was on a ventilator; and a story about a COVID-19 naysayer who regretted not having gotten the vaccine after suffering a major health event (see Figures 2 and 3).

**Community Connections**

A second theme was more positive framing focused on family and community connections. Participants emphasized the need to make people feel like part of a community. They offered

images of, for example, gatherings among extended family and singing spirituals together. One approach to conveying this theme was to highlight that we all face similar struggles and that getting the vaccination shows concern about and for the community. A second focused on the idea of a return to normalcy and being able to live one's best life. For example, getting vaccinated meant that a person could move back out into the world, connect with friends and family members, and participate once again in important events, such as seeing a sister graduate from law school.

### ***Addressing Vaccine Hesitancy***

Several participants acknowledged that discussing COVID-19 vaccines was challenging, either for themselves or for their families. One participant indicated that people had strong opinions on both sides—both for and against the vaccine. She and others intimated that the vaccine should not be politicized or weaponized on either side and that the more aggressive the stance, the more resistant people will be. To be most effective, we must show patience, provide different perspectives, and be respectful in conversations with reluctant family members. Several ideas emerged during brainstorming, including showing someone who did not want the vaccine but whose mind was changed by an event or person, someone who feels it is not OK to mandate but still got the vaccine themselves, or someone afraid of needles but who was helped by a friend to get through it.

### ***Countering Vaccine Misinformation***

A final theme focused on the need to debunk vaccine rumors and misinformation. Participants noted that some members of their community raised questions about what is in the vaccine and whether the vaccine could change your DNA. They also noted deep-seated mistrust of the medical establishment and concerns that previous vaccines had been tested on Black people as well as lingering resistance because you can still get COVID-19 if vaccinated. Final videos reflected these concerns, as well as various counterarguments participants offered. For example, one participant equated motivations for prevention of COVID-19 and pregnancy by pointing out that we still encourage condom use even though “you can still get pregnant.” Some highlighted the persuasiveness of using celebrities to discuss and combat vaccine misinformation or of featuring positive experiences from breastfeeding moms who got the vaccine.

## **RQ 2: Issues Related to Production Techniques and Techniques Used**

### ***Production Technique Issues and Process***

In the first 2 workshops, our video experts reviewed basic technical procedures and how-to tips on filming and editing

videos. Digital storytelling participants were invited to consider several technical aspects as they planned their videos. They included (1) venues, (2) how to shoot their video, (3) who the “speaker” for the video is, and (4) any unique creative elements they might use.

Regarding brainstorming venues, participants seemed comfortable envisioning appropriate backdrops. In the breakout sessions, participants suggested a diverse set of backgrounds, including a public park or church, the Birmingham Civil Rights Institute, or sitting on a porch with friends.

Regarding how to record videos, during breakout sessions, a few participants described previous experience shooting and editing videos, but most had not had much experience with that and expressed concern about being able to produce videos. Our video experts reminded them to be gentle with themselves, shoot more, and edit out. Participants were also encouraged to use the technology they had (cell phone, tablet, or computer) rather than obtaining special equipment or software.

Regarding speakers, as they brainstormed initial story ideas, some participants seemed comfortable placing themselves at the center of their stories, while others imagined their story delivered by others (for example, a trusted elder or stand-in actors).

For unique creative elements, ideas generated during the breakout sessions included music or dramatic news headlines to set a somber tone and then shifting scenery, speaker, or music to move viewers' emotions from dark or anxious to hopeful. Other ideas included depiction of an instant messaging conversation between friends who were supporting each other to resist vaccine-related misinformation.

### ***Technical Approaches Used in Final Digital Stories***

Summarized in Table 1 are each of the 11 videos. Most of the videos (n=7) incorporated music at some point during the story to convey emotions or to serve as a backdrop. Only a handful of stories utilized subtitles (n=3), most notably used in a video that depicted a text message-based conversation with peers about COVID-19 vaccines. Unique creative elements were seen in every digital story, which shows the individual creativity of each creator using the skills taught during the workshops. The elements included, but were not limited to, continuous narration (video 1), animations to add visual interest (videos 2 and 8), multimedia platform usage with transitions (video 4), and incorporation of news or social media headlines and addressing misinformation (videos 9 and 11).

**Table 1.** Digital story loglines and content from digital storytelling workshop participants (N=11).

Video	Logline	Sex	Age (years)	Theme(s)	Music	Subtitles	Unique creative elements
1	A young student somberly reflects on why she chose to get the COVID-19 vaccine.	F <sup>a</sup>	21	1. Community connections 2. Addressing vaccine hesitancy	Yes	No	The video is narrated with overlaid audio and edited together with simple transitions. There is a good balance of narrative footage (where she speaks directly into the camera) and cutaway footage (contextual video clips), which provide a clear digital story. Ambient music played softly in the background to strengthen the reflective tone of the video.
2	When a young woman is not receiving support from her family about the COVID-19 vaccine, her close friends are there to give her the validation she needs.	F	21	1. COVID-19 vulnerability 2. Community connections 3. Countering vaccine misinformation	No	Yes	The video begins with a short title and introductory slide, then quickly fades to a simulated screen-capture of someone typing and sending a text message (recreated via animation). The video features 2 text message threads. The first conversation is between her and some unsupportive family members, and the second is between her and a group of supportive friends. Each is clearly demarcated with different message colors, and the conversations are separated by a narrative slide with bold yellow text on a black background. The video ends with a closing slide that is styled in the same way for visual continuity. The video effectively uses sound effects to replicate the experience of an active text conversation.
3	When a young woman is hesitant about receiving the COVID-19 vaccine, the possibility of missing out on major family moments motivates her to decide.	F	26	1. COVID-19 vulnerability 2. Community connections 3. Addressing vaccine hesitancy	Yes	Yes	The video begins with a blank screen and the sounds of a person getting into their car/turning on the radio. Footage from driving to a pharmacy is shot from the car dashboard (street view). It also captures her walking into the pharmacy for her vaccination appointment. This “in-car” perspective shows up again in footage of her family celebrating her sister’s graduation from their car. Throughout the video, black slides with white text fade onto the screen to move the narrative along. Although the first part of her video does not include music (just natural background sounds from her car), in the second half of the video, she overlays upbeat music on a slideshow of pictures and video clips capturing the postgraduation festivities. The music contrast conveys the contrast of her initial feelings of anxiety and hesitancy with her eventual excitement and relief.
4	A young woman experiences a date-gone-wrong when talking about COVID-19 misinformation with her date.	F	24	1. Community connections 2. Countering vaccine misinformation	No	No	The video begins with a Zoom-esque call between her and her 2 friends. All 3 videos are displayed on the screen (gallery style), and you can see that they are filming from different locations. They greet each other, and her friends are curious to know more about a date she recently went on. After sharing that it did not go well, she starts to describe the last moments of the date. She begins, “ <i>Pretty much it went like this...</i> ”—the video uses a jump cut to take us directly to a scene/conversation from her date. The dialogue scenes are edited together using a mix of wide and closeup shots. The wide-shot framed her and her date as they sit on a couch. Tighter, over-the-shoulder shots focused in on her date and his perspective on COVID-19. The date ends with her wishing him well and encouraging him to change his views on COVID-19 vaccination. The video jumps back to the video call between the sisters for the final wrap up. They laugh together and end their video call.

Video	Logline	Sex	Age (years)	Theme(s)	Music	Subtitles	Unique creative elements
5	A young mother reflects on her most isolating and impactful moments from the beginning of the pandemic.	F	26	<ol style="list-style-type: none"> <li>COVID-19 vulnerability</li> <li>Community connections</li> </ol>	Yes	No	The video begins with dramatic music and an all-caps title slide that reads "2020 THE YEAR OF COVID-19." The words are bright red and are on a black background. As the burst of music fades, the video quickly transitions to footage of her speaking directly to the camera. She filmed outdoors, her lighting was even, and the audio was clear.
6	A young woman wants to encourage people to get the COVID-19 vaccine with hopes of getting our lives back before the pandemic.	F	23	<ol style="list-style-type: none"> <li>COVID-19 vulnerability</li> <li>Community connections</li> </ol>	Yes	No	The video opens with a title slide and fades to footage of a woman sitting on a sofa—staring straight ahead. As the strings play the intro of a well-known, somber, soul music song, the video zooms in on her emotionless expression. The clip has a blue monotone filter, and the year 2020 is fixed to the top left corner of the screen. As the camera inches closer, she stretches out her TV remote to signal that she is changing the channel as the music plays for a few bars. The screen transitions to black and uses a quick <i>iris close</i> transition to mimic turning off a TV screen. At the same time, she uses the <i>turntable rewind</i> sound effect to signal a shift in the narrative. The screen is black, and the date 2021 appears in the top left corner. An uplifting, high-tempo, R&B song begins playing. As the music builds, you see footage of her entering a pharmacy for her vaccine appointment. The visual style of the video changes from this point on, showing full color and the edits timed to match the upbeat rhythm of the song. In some areas, she plays with speeding up the footage or removing frames (stop-motion aesthetic) to highlight special moments (eg, showing off her vaccination Band-Aid, hugging family). The video continues in this style as she drives to her family's home to play Monopoly. As the footage continues and the music plays softly, she adds a voiceover narration. She encourages people to get vaccinated so we can "safely get back to enjoying life." The video ends with an end card slide and #ToughTalks.
7	A classroom engages in insightful conversation on the importance of the COVID-19 vaccine.	F	26	<ol style="list-style-type: none"> <li>Community connections</li> <li>Addressing vaccine hesitancy</li> <li>Countering vaccine misinformation</li> </ol>	No	No	The video begins with a stock photo of a smiling teacher standing in front of a chalkboard. The video continues in this aesthetic and combines voiceovers with stock photos to visualize a conversation between a teacher and inquisitive students. The transitions cut from photo to photo, and the photos correspond with the character voice that is speaking (characters: teacher and students Ashely, David, Veronica, Lindsey).
8	COVID-19 superwoman shares tips and information about the COVID-19 vaccine.	F	19	<ol style="list-style-type: none"> <li>COVID-19 vulnerability</li> <li>Community connections</li> </ol>	Yes	No	The video is informational and presented as a series of animated slides. Some of the slide graphics include simple animations to add visual interest. The music is ambient and does not distract from the content.



Video	Logline	Sex	Age (years)	Theme(s)	Music	Subtitles	Unique creative elements
9	A young woman addresses myths about the COVID-19 vaccine.	F	19	1. Addressing vaccine hesitancy 2. Countering vaccine misinformation	Yes	No	The video begins with dramatic, high-energy music playing in the background and a screenshot of a tweet from a well-known rapper who has expressed COVID-19 vaccine hesitancy through social media. After a few seconds, the clip and music abruptly stop, and the video cuts to a CNN news clip addressing the misinformation in the rapper's tweet. With the context and format established (present misinformation, and address it head on), her video transitions to her speaking directly to the camera. She continues to address common rumors and misinformation surrounding COVID-19. She used graphics and supporting screenshots as she told the story.
10	A young woman reflects on her experience with missing loved ones and how getting vaccinated brought back the connections she longed for.	F	24	1. COVID-19 vulnerability 2. Community connections	No	No	The video is a personal account of a young woman's experience with COVID-19. The video was filmed in one continuous take, and she shared her story by speaking directly into the camera. Her background was styled with neutral décor to add visual interest but not distract from herself. Her lighting was in front of her and even cast her face, and her audio was consistent and clear.
11	A young woman shares how her mother being a nurse influenced her decision to get the COVID-19 vaccine.	F	22	1. COVID-19 vulnerability 2. Community connections	Yes	Yes	The video is set in New York. It opens with a slide that reads "My COVID-19 experience was so much different than anything I've heard" as a somber, hopeful piano track plays in the background. The video transitions between news clip headlines, photos of empty city streets, and exhausted health care workers. While the music continues in the background, a news clip featuring an interview with her mother, a nurse, plays. Throughout the video, black slides with white text fade onto the screen to move the narrative along.

<sup>a</sup>F: female.

### RQ 3: Overall Assessment of the Digital Storytelling Workshop Process

In the quantitative workshop feedback, participants identified both professional and personal value from learning how to tell digital stories. Based on evaluation forms and informal feedback at the end of the third workshop, participants appreciated learning technical skills for both storytelling and videography (Figure 4).

All participants agreed that their time was well spent, with almost three-quarters (8/11, 73%) strongly agreeing. The majority (9/11, 82%) strongly agreed that the digital storytelling workshops were like what had been explained to them; 10 of 11 agreed (n=5) or strongly agreed (n=5) that they had some ideas about what story to tell by the end of the first workshops, and most (8/11, 73%) strongly agreed they had narrowed down their ideas by workshop two. Of the 11 participants, 9 felt they would very likely (n=6) or likely (n=3) use these digital storytelling techniques for personal use in the future, and even more were very likely (n=7) to use the techniques for professional use. Most participants (6/11, 55%) spent between 3 hours to 4 hours shooting their videos. Just under one-half (4/11, 36%) spent about 3 hours to 4 hours editing their video.

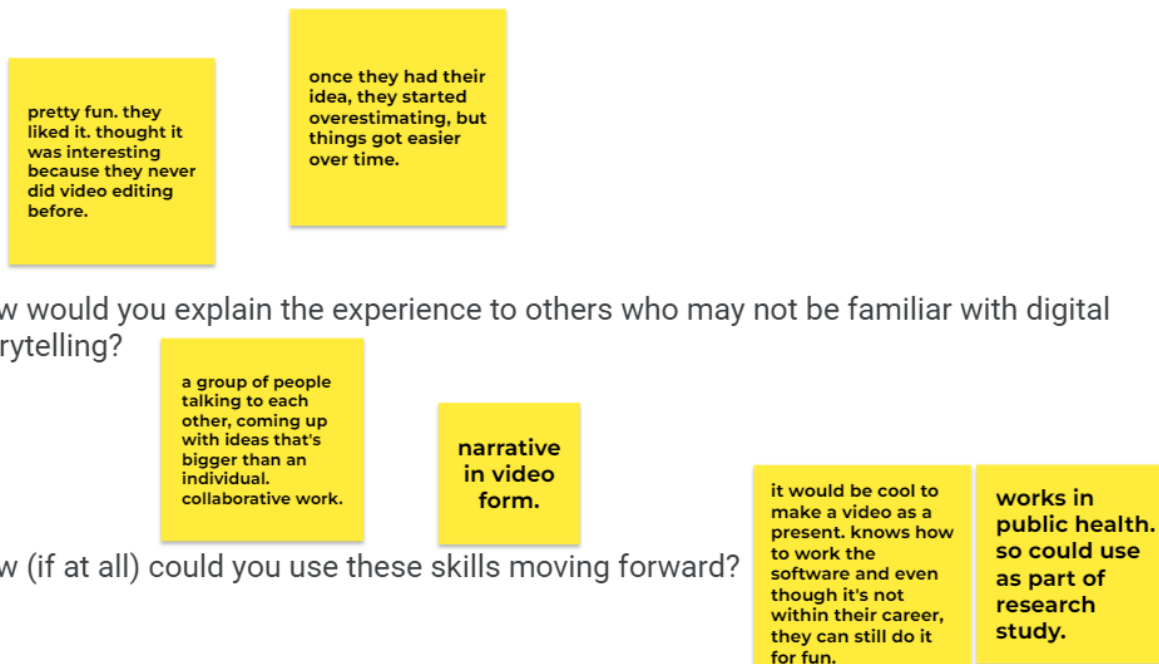
All participants said they would be interested in being contacted in the future for a related Tough-Talks: COVID or other research activity.

When asked about how they might explain what they learned about digital storytelling to others, one participant described it as "sort of like TikTok but with really personal stories," while another described it as learning "how to put your ideas and experiences out there for others—not just for your family." One participant liked learning how to frame ideas more clearly and script them to be purposeful. Several participants appreciated the introduction to video editing and acknowledged that these (new-found, for many) digital storytelling skills could help with classes, presentations, job searches, and work-related activities. Participants came to these workshops with different motivations and expectations. Several joined out of curiosity or to accept a challenge; some worried that their personal story would not measure up. All who completed the workshop series shared positive remarks at the end. On the personal side, one participant reflected that she now felt prepared to share not just digital stories but also personal stories. Many of the participants felt validated or even empowered from hearing the diverse perspectives about why their peers had chosen to get vaccinated.

**Figure 4.** Digital storytelling workshop feedback from participants during the final session.

Questions for participants:

How do you feel about the workshops, now that you have completed them?



## Discussion

The purpose of this research was to describe the use of digital storytelling methods to empower young Black adults in the South who continue to experience COVID-19 disparities by documenting what type of stories were told and how they emerged, as well as the way in which these individuals felt about the process.

### Principal Findings and Application to Ongoing Study Activities

Overall, we observed high satisfaction with workshop participation, and every participant reported professional benefits to digital storytelling training and that workshops were a good use of their time. This is a unique innovation of our approach, which, when compared with PhotoVoice, may offer more professional as well as personal relevance to young people who often share about their lives in social media formats but may not be empowered to describe their views on COVID-19 vaccine decision-making.

The videos produced have potential to resonate with peers because they reflect the unique voices of participants and may invoke shared experiences. Videos also addressed risk factors for Black communities (such as vaccine misinformation) and highlighted cultural strengths (such as community connections). Stories were conveyed using a variety of techniques, most of which were more sophisticated than simply capturing video on one's phone.

In future digital storytelling workshops, our team would increase the interactive nature of the sessions, such as a devoted workshop session where draft versions receive hands-on guidance from our content and technical team leaders. Another

improvement for future workshops would be a brief inventory about familiarity with digital storytelling that could be shared with participants ahead of workshops to gauge their existing familiarity with and interest in social media engagement and digital storytelling. This type of inventory would be adapted to other health topics and also other populations to better tailor workshop content. Lastly, given the engagement of our youth advisors throughout all aspects of the study, youth advisors could be invited to attend digital storytelling workshops, create digital stories themselves, or lead aspects of the training that may yield professional benefit to them like that expressed by workshop participants.

### Comparison With Prior Work

Our approach, guided by CBPR principles, was conducive to centering the needs and voices of communities experiencing marginalization, such as young adults of color. Compared with common CBPR methods such as PhotoVoice, digital stories similarly focus on empowering communities, with added innovative ability to incorporate mixed media and conduciveness to dissemination through social media. Given the ubiquity of social media, digital storytelling holds powerful potential to combat COVID-19 misinformation when told from the vantage point of young Black adults to peers.

As the larger Tough Talks-COVID study gets underway, we continue to consider ways to address the ethical standards of CBPR described earlier: (1) how to achieve a true community-driven agenda, (2) the tension of insiders versus outsiders, (3) the reality of participation limitations, and (4) shared ownership and dissemination of findings to drive action [20]. We have worked to achieve a true community-driven agenda with the formation and maintenance of youth and expert advisory board members who are from and work in the

communities in the South where our study is based. Our experts include community advocates, bioethicists, pastors, and vaccine scientists and is 93% Black (youth advisory board is 100% Black). To date, our advisory board members have edited and tested our baseline surveys, informed the scripts for informational videos that are being built into the Tough Talks-COVID intervention, and were featured judges for a showcase where our digital storytelling workshop participants competed for additional cash prizes from incentives previously mentioned.

Next, we navigate insider-outsider tension by having several team members who identify as young people of color (3 of the 5 digital storytelling team members), and most of our larger Tough Talks-COVID team live and work in the 3 regions where the study is being conducted with a longstanding history of community-engaged work. Third, while digital storytelling workshop participants responded positively to engagement methods such as the audience-specific emotion cards, future study activities can do more engaged work via in-person activities that were not feasible for us considering the COVID-19 global pandemic. Lastly, shared ownership and dissemination include that, upon progression of the study, participants will be invited to share their digital stories on their own social media. We will also recontact workshop participants, since all consented to further study contact, to invite them along with youth advisors to conduct their own content analysis of their digital stories and provide training to them on this along with how to craft a scholarly manuscript that they will coauthor.

### Limitations

The present findings are subject to limitations. Our digital storytelling workshops (and the broader Tough Talks-COVID study) focus on young Black adults in the South and was limited in size. Although there are theoretical, epidemiological, and empirical reasons for doing so, the exclusive focus may limit the generalizability of our findings. We found that participants who self-selected into workshops were 100% vaccine-accepting despite our invitation being open also to those who are vaccine-ambivalent. Greater variability may have yielded

different stories and themes with greater potential to be useful to the Tough Talks-COVID app. The videos produced had little male representation, reflecting the demographics of the larger survey sample from which the participants were selected. Additionally, 2 men dropped out before completing the workshops. Our workshops did not emphasize diversity in ethnicity (eg, Hispanic/Latinx, Afro-Caribbean), sexual orientation and gender identity, or disability status. Finally, although we describe themes and features of stories and participants' perceptions of workshop experiences, these findings do not speak to the effectiveness of videos, individually, on influencing attitudes and behaviors or in combatting misinformation. Experimental research is needed to assess that kind of impact. Selected videos will be assessed as a component in a larger digital health intervention by the Tough Talks-COVID study.

### Conclusions

Despite limitations, to our knowledge, our Tough Talks-COVID study is one of the first to incorporate digital storytelling as a central component to a digital health intervention and the only one to do so with exclusive focus on young Black adults. We feel this is appropriate, given the deeply rooted mistrust of many medical establishments including those who have produced the COVID-19 vaccine, which has often taken the form of misinformation spreading among young adults, including through social media. Our emphasis on digital storytelling was shown to be highly acceptable, and future activities with this approach in our study will engage more diversity within this population to contribute additional digital stories and invite young people to disseminate these stories on their own social media as a powerful peer-driven approach to combat COVID-19 misinformation online. Similar approaches, including careful consideration of the ethical challenges of CBPR approaches, are applicable to other populations experiencing both COVID-19 inequities and marginalization, such as other age demographics and people of color. More research is needed to disseminate this innovative approach and leverage its profound potential to uplift the voices and needs of communities experiencing marginalization and COVID-19 inequities.

### Acknowledgments

Our research was supported by funding from the National Institutes for Minority Health and Health Disparities (1R01MD016834).

### Authors' Contributions

AM-B led all digital storytelling activities and co-designed the curriculum with MLC, EET, JW, and DSJ. CT led study recruitment and provided feedback on digital storytelling approaches as part of coordination of the larger Tough Talks: COVID study. HB and LH-W are principal investigators and site leads for the larger study and provided oversight to all study components including digital storytelling workshops.

### Conflicts of Interest

None declared.

### References

1. Yancy CW. COVID-19 and African Americans. *JAMA* 2020 May 19;323(19):1891-1892. [doi: [10.1001/jama.2020.6548](https://doi.org/10.1001/jama.2020.6548)] [Medline: [32293639](https://pubmed.ncbi.nlm.nih.gov/32293639/)]

2. Raifman MA, Raifman JR. Disparities in the population at risk of severe illness from COVID-19 by race/ethnicity and income. *Am J Prev Med* 2020 Jul;59(1):137-139 [FREE Full text] [doi: [10.1016/j.amepre.2020.04.003](https://doi.org/10.1016/j.amepre.2020.04.003)] [Medline: [32430225](https://pubmed.ncbi.nlm.nih.gov/32430225/)]
3. Wakeel F, Njoku A. Application of the weathering framework: intersection of racism, stigma, and COVID-19 as a stressful life event among African Americans. *Healthcare (Basel)* 2021 Feb 02;9(2):145 [FREE Full text] [doi: [10.3390/healthcare9020145](https://doi.org/10.3390/healthcare9020145)] [Medline: [33540498](https://pubmed.ncbi.nlm.nih.gov/33540498/)]
4. Riggle EDB, Drabble LA, Bochicchio LA, Wootton AR, Veldhuis CB, Munroe C, et al. Experiences of the COVID-19 pandemic Among African American, Latinx, and White sexual minority women: a descriptive phenomenological study. *Psychol Sex Orientat Gend Divers* 2021 Jun;8(2):145-158 [FREE Full text] [doi: [10.1037/sgd0000510](https://doi.org/10.1037/sgd0000510)] [Medline: [34746332](https://pubmed.ncbi.nlm.nih.gov/34746332/)]
5. Czeisler ME, 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](https://doi.org/10.15585/mmwr.mm6932a1)] [Medline: [32790653](https://pubmed.ncbi.nlm.nih.gov/32790653/)]
6. Fortuna LR, Tolou-Shams M, Robles-Ramamurthy B, Porche MV. Inequity and the disproportionate impact of COVID-19 on communities of color in the United States: The need for a trauma-informed social justice response. *Psychol Trauma* 2020 Jul;12(5):443-445 [FREE Full text] [doi: [10.1037/tra0000889](https://doi.org/10.1037/tra0000889)] [Medline: [32478545](https://pubmed.ncbi.nlm.nih.gov/32478545/)]
7. Bogart L, Ojikutu B, Tyagi K, Klein DJ, Mutchler MG, Dong L, et al. COVID-19 related medical mistrust, health impacts, and potential vaccine hesitancy among Black Americans living with HIV. *J Acquir Immune Defic Syndr* 2021 Feb 01;86(2):200-207 [FREE Full text] [doi: [10.1097/QAI.0000000000002570](https://doi.org/10.1097/QAI.0000000000002570)] [Medline: [33196555](https://pubmed.ncbi.nlm.nih.gov/33196555/)]
8. Eaton LA, Driffin DD, Kegler C, Smith H, Conway-Washington C, White D, et al. The role of stigma and medical mistrust in the routine health care engagement of Black Men who have sex with men. *Am J Public Health* 2015 Feb;105(2):e75-e82. [doi: [10.2105/ajph.2014.302322](https://doi.org/10.2105/ajph.2014.302322)]
9. Collins-Dexter B. *Canaries in the Coal Mine: COVID-19 Misinformation and Black Communities*. Shorenstein Center on Media, Politics and Public Policy. 2020. URL: <https://shorensteincenter.org/wp-content/uploads/2020/06/Canaries-in-the-Coal-Mine-Shorenstein-Center-June-2020.pdf> [accessed 2022-09-02]
10. Kricorian K, Turner K. COVID-19 vaccine acceptance and beliefs among Black and Hispanic Americans. *PLoS One* 2021 Aug 24;16(8):e0256122 [FREE Full text] [doi: [10.1371/journal.pone.0256122](https://doi.org/10.1371/journal.pone.0256122)] [Medline: [34428216](https://pubmed.ncbi.nlm.nih.gov/34428216/)]
11. Auxier B, Anderson M. *Social Media Use in 2021*. Pew Research Center. 2021 Apr 07. URL: <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/> [accessed 2022-09-02]
12. Miller GH, Marquez-Velarde G, Williams AA, Keith VM. Discrimination and Black social media use: sites of oppression and expression. *Sociology of Race and Ethnicity* 2020 Sep 02;7(2):247-263. [doi: [10.1177/2332649220948179](https://doi.org/10.1177/2332649220948179)]
13. Kreuter MW, Green MC, Cappella JN, Slater MD, Wise ME, Storey D, et al. Narrative communication in cancer prevention and control: A framework to guide research and application. *Ann Behav Med* 2007 Sep;33(3):221-235. [doi: [10.1007/bf02879904](https://doi.org/10.1007/bf02879904)]
14. de Jager A, Fogarty A, Tewson A, Lenette C, Boydell K. Digital storytelling in research: a systematic review. *TQR* 2017 Oct 2;2548-2582. [doi: [10.46743/2160-3715/2017.2970](https://doi.org/10.46743/2160-3715/2017.2970)]
15. Sood S, Riley A, Alarcon K. Entertainment-education and health and risk messaging. *Oxford Research Encyclopedia of Communication* 2017:1 [FREE Full text] [doi: [10.1093/acrefore/9780190228613.013.245](https://doi.org/10.1093/acrefore/9780190228613.013.245)]
16. Moyer-Gusé E. Toward a theory of entertainment persuasion: explaining the persuasive effects of entertainment-education messages. *Communication Theory* 2008;18(3):407-425. [doi: [10.1111/j.1468-2885.2008.00328.x](https://doi.org/10.1111/j.1468-2885.2008.00328.x)]
17. Collins SE, Clifasefi SL, Stanton J, The LEAP Advisory Board, Straits KJE, Gil-Kashiwabara E, et al. Community-based participatory research (CBPR): Towards equitable involvement of community in psychology research. *Am Psychol* 2018 Oct;73(7):884-898 [FREE Full text] [doi: [10.1037/amp000167](https://doi.org/10.1037/amp000167)] [Medline: [29355352](https://pubmed.ncbi.nlm.nih.gov/29355352/)]
18. Wallerstein NB, Duran B. Using community-based participatory research to address health disparities. *Health Promot Pract* 2006 Jul;7(3):312-323. [doi: [10.1177/1524839906289376](https://doi.org/10.1177/1524839906289376)] [Medline: [16760238](https://pubmed.ncbi.nlm.nih.gov/16760238/)]
19. Lightfoot E, McCleary JS, Lum T. Asset mapping as a research tool for community-based participatory research in social work. *Social Work Research* 2014 Mar 13;38(1):59-64. [doi: [10.1093/swr/svu001](https://doi.org/10.1093/swr/svu001)]
20. Minkler M. Ethical challenges for the "outside" researcher in community-based participatory research. *Health Educ Behav* 2004 Dec 01;31(6):684-697. [doi: [10.1177/1090198104269566](https://doi.org/10.1177/1090198104269566)] [Medline: [15539542](https://pubmed.ncbi.nlm.nih.gov/15539542/)]
21. Stoner MCD, Browne E, Tweedy D, Pettifor A, Maragh-Bass AC, Toval C, et al. Exploring motivations for COVID-19 vaccination among Black young adults in three Southern states: a cross sectional study between September to November 2021. *JMIR Form Res* 2022 Aug 15;6(9):e39144. [doi: [10.2196/39144](https://doi.org/10.2196/39144)] [Medline: [35969516](https://pubmed.ncbi.nlm.nih.gov/35969516/)]

## Abbreviations

- BIPOC:** Black, Indigenous, and other People of Color
- CBPR:** community-based participatory research
- HIPAA:** Health Insurance Portability and Accountability Act
- RCT:** randomized controlled trial
- RQ:** research question

*Edited by A Mavragani; submitted 01.04.22; peer-reviewed by J Bailey, O Rahaman; comments to author 17.05.22; revised version received 15.08.22; accepted 16.08.22; published 26.09.22.*

*Please cite as:*

*Maragh-Bass A, Comello ML, Tolley EE, Stevens Jr D, Wilson J, Toval C, Budhwani H, Hightow-Weidman L*

*Digital Storytelling Methods to Empower Young Black Adults in COVID-19 Vaccination Decision-Making: Feasibility Study and Demonstration*

*JMIR Form Res 2022;6(9):e38070*

*URL: <https://formative.jmir.org/2022/9/e38070>*

*doi: [10.2196/38070](https://doi.org/10.2196/38070)*

*PMID: [36155984](https://pubmed.ncbi.nlm.nih.gov/36155984/)*

©Allysha Maragh-Bass, Maria Leonora Comello, Elizabeth Ellen Tolley, Darrell Stevens Jr, Jade Wilson, Christina Toval, Henna Budhwani, Lisa Hightow-Weidman. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 26.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Corrigenda and Addenda

# Correction: Using the Transformative Storytelling Technique to Generate Empowering Narratives for Informal Caregivers: Semistructured Interviews, Thematic Analysis, and Method Demonstration

Milica Petrovic<sup>1\*</sup>, MSc; Silvia Bonanno<sup>2\*</sup>, MSc; Marta Landoni<sup>3</sup>, PhD; Chiara Ionio<sup>3</sup>, PhD; Mariët Hagedoorn<sup>4\*</sup>, PhD; Andrea Gaggioli<sup>5,6\*</sup>, PhD

<sup>1</sup>ExperienceLab, Università Cattolica del Sacro Cuore, Milan, Italy

<sup>2</sup>Università Cattolica del Sacro Cuore, Brescia, Italy

<sup>3</sup>Department of Psychology, Centro di Ricerca sulle Dinamiche Evolutive ed Educative, Università Cattolica del Sacro Cuore, Milan, Italy

<sup>4</sup>Department of Health Psychology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

<sup>5</sup>Research Center in Communication Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

<sup>6</sup>IRCCS, Istituto Auxologico Italiano, Milan, Italy

\*these authors contributed equally

**Corresponding Author:**

Milica Petrovic, MSc

ExperienceLab

Università Cattolica del Sacro Cuore

Largo A Gemmeli 1

Milan, 20123

Italy

Phone: 39 3497366542

Email: [milica.petrovic@unicatt.it](mailto:milica.petrovic@unicatt.it)

**Related Article:**

Correction of: <https://formative.jmir.org/2022/8/e36405>

(*JMIR Form Res* 2022;6(9):e42323) doi:[10.2196/42323](https://doi.org/10.2196/42323)

In “Using the Transformative Storytelling Technique to Generate Empowering Narratives for Informal Caregivers: Semistructured Interviews, Thematic Analysis, and Method Demonstration” (*JMIR Form Res* 2022;6(8):e36405), one error was noted.

In the originally published paper, Affiliation 6 for author Andrea Gaggioli appeared as follows:

*Applied Technology for Neuro-Psychology Lab,  
Istituto Auxologico Italiano, Milan, Italy*

In the corrected version of the paper, Affiliation 6 has been revised as follows:

*IRCCS, Istituto Auxologico Italiano, Milan, Italy*

The correction will appear in the online version of the paper on the JMIR Publications website on September 14, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

*Submitted 31.08.22; this is a non-peer-reviewed article; accepted 31.08.22; published 14.09.22.*

*Please cite as:*

*Petrovic M, Bonanno S, Landoni M, Ionio C, Hagedoorn M, Gaggioli A*

*Correction: Using the Transformative Storytelling Technique to Generate Empowering Narratives for Informal Caregivers: Semistructured Interviews, Thematic Analysis, and Method Demonstration*

*JMIR Form Res 2022;6(9):e42323*

*URL: <https://formative.jmir.org/2022/9/e42323>*

*doi: [10.2196/42323](https://doi.org/10.2196/42323)*

*PMID: [36103710](https://pubmed.ncbi.nlm.nih.gov/36103710/)*

©Milica Petrovic, Silvia Bonanno, Marta Landoni, Chiara Ionio, Mariët Hagedoorn, Andrea Gaggioli. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 14.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Discovering Long COVID Symptom Patterns: Association Rule Mining and Sentiment Analysis in Social Media Tweets

Surani Matharaarachchi<sup>1</sup>, BSc, MSc; Mike Domaratzki<sup>2</sup>, BMATH, MMATH, PhD; Alan Katz<sup>3</sup>, MBChB, MSc, CCFP; Saman Muthukumarana<sup>1</sup>, BSc, MSc, PhD

<sup>1</sup>Department of Statistics, University of Manitoba, Winnipeg, MB, Canada

<sup>2</sup>Department of Computer Science, Western University, London, ON, Canada

<sup>3</sup>Departments of Community Health Sciences and Family Medicine, University of Manitoba, Winnipeg, MB, Canada

**Corresponding Author:**

Surani Matharaarachchi, BSc, MSc

Department of Statistics

University of Manitoba

66 Chancellors Cir

Winnipeg, MB, R3T 2N2

Canada

Phone: 1 431 338 5077

Email: [matharas@myumanitoba.ca](mailto:matharas@myumanitoba.ca)

## Abstract

**Background:** The COVID-19 pandemic is a substantial public health crisis that negatively affects human health and well-being. As a result of being infected with the coronavirus, patients can experience long-term health effects called long COVID syndrome. Multiple symptoms characterize this syndrome, and it is crucial to identify these symptoms as they may negatively impact patients' day-to-day lives. Breathlessness, fatigue, and brain fog are the 3 most common continuing and debilitating symptoms that patients with long COVID have reported, often months after the onset of COVID-19.

**Objective:** This study aimed to understand the patterns and behavior of long COVID symptoms reported by patients on the Twitter social media platform, which is vital to improving our understanding of long COVID.

**Methods:** Long COVID-related Twitter data were collected from May 1, 2020, to December 31, 2021. We used association rule mining techniques to identify frequent symptoms and establish relationships between symptoms among patients with long COVID in Twitter social media discussions. The highest confidence level-based detection was used to determine the most significant rules with 10% minimum confidence and 0.01% minimum support with a positive lift.

**Results:** Among the 30,327 tweets included in our study, the most frequent symptoms were brain fog (n=7812, 25.8%), fatigue (n=5284, 17.4%), breathing/lung issues (n=4750, 15.7%), heart issues (n=2900, 9.6%), flu symptoms (n=2824, 9.3%), depression (n=2256, 7.4%) and general pains (n=1786, 5.9%). Loss of smell and taste, cold, cough, chest pain, fever, headache, and arm pain emerged in 1.6% (n=474) to 5.3% (n=1616) of patients with long COVID. Furthermore, the highest confidence level-based detection successfully demonstrates the potential of association analysis and the Apriori algorithm to establish patterns to explore 57 meaningful relationship rules among long COVID symptoms. The strongest relationship revealed that patients with lung/breathing problems and loss of taste are likely to have a loss of smell with 77% confidence.

**Conclusions:** There are very active social media discussions that could support the growing understanding of COVID-19 and its long-term impact. These discussions enable a potential field of research to analyze the behavior of long COVID syndrome. Exploratory data analysis using natural language processing methods revealed the symptoms and medical conditions related to long COVID discussions on the Twitter social media platform. Using Apriori algorithm-based association rules, we determined interesting and meaningful relationships between symptoms.

(JMIR Form Res 2022;6(9):e37984) doi:[10.2196/37984](https://doi.org/10.2196/37984)

**KEYWORDS**

COVID-19; long COVID symptoms; social media analysis; association rule mining; bigram analysis; natural language processing; Twitter; content analysis; data mining; infodemiology; health information



## Introduction

COVID-19, a transmissible disease caused by the SARS-CoV-2 virus, has become a substantial public health crisis that negatively affects people's health and well-being. Most people with COVID-19 recover entirely within weeks. However, some people still experience symptoms after their initial recovery, even those who had mild symptoms with their initial infection. Others develop new symptoms related to their COVID-19 illness. These people sometimes describe themselves as "long haulers" [1]. This syndrome has been described as post-COVID-19 or "long COVID-19" [2]. It is crucial to identify these symptoms as they may negatively impact the day-to-day lives of those with this affliction for a substantially long period. Breathlessness, fatigue, and brain fog are the 3 most common symptoms that patients with long COVID have consistently reported, often months after the onset of the COVID-19 disease [3,4].

Social media has become a substantial part of our lives. People use it to connect with others and share their thoughts, emotions, and experiences about any current topic, often without revealing their identity [5]. Hence, social media platforms such as Twitter, Facebook, and Instagram have had a massive impact on society and gained considerable research attention [6]. There are extensive chains of discussions on these platforms about the long COVID syndrome (LCS). Thus, analyzing social media conversations of long COVID-related patients from various sources provides an opportunity to understand the relationship between symptoms and their consequences.

According to the World Health Organization clinical case definition [4], "long COVID-19 condition occurs in individuals with a history of probable or confirmed SARS CoV2 infection, usually three months from the onset of COVID-19, with symptoms that last for at least two months." Health-related organizations, such as the Mayo Clinic [7], National Health Service [8], Centers for Disease Control and Prevention [9], and World Health Organization [4], have identified different lists of symptoms related to long COVID, and a summary is presented below. The results of recent research that has been conducted using self-reported long COVID symptoms on Twitter are presented [10]. These symptoms were identified by manually reading 165 tweets from July 20, 2020, to July 29, 2020.

There are, however, very active ongoing social media discussions that could support the growing understanding of the illness and its long-term impact. These discussions provide the opportunity to access publicly available data from multiple individuals on Twitter to analyze long COVID symptoms. However, manually discovering the knowledge in a large volume of unstructured texts is increasingly problematic. Hence, automated natural language processing (NLP) methods have been introduced to do this task effectively and accurately [11,12]. The importance of using NLP methods in health sciences has been increasingly recognized over recent years [13]. Even though previous research has identified a list of symptoms, extracting and grouping the symptoms discussed in tweets into categories makes them easier to analyze and find the relationships between the most common symptoms.

Manually handling this task is challenging; thus, NLP tools represent an opportunity to extract hidden information from unstructured Twitter text data.

Association rules are considered a useful tool as they offer the possibility to conduct intelligent diagnoses, extract invaluable information, and build important knowledge quickly and automatically while identifying relationships within and between variables [14]. Thus, we used association rule mining (ARM) to discover relationships among long COVID symptoms based on the symptoms revealed from text data [15]. The mining process of association rules requires setting a minimum confidence and support threshold to describe association rules that are meaningful. Therefore, we identified association rules based on a minimum confidence threshold of 10% and a minimum support threshold of 0.1%.

This study sought to achieve the 2 goals. The first goal was to identify the symptoms and medical conditions related to long COVID that were discussed on the Twitter social media platform. The second goal was to determine the patterns of symptoms and their associations. By accomplishing these objectives, this work will ultimately help physicians identify the behavior of the patients with long COVID. This paper provides new ideas for symptom mining and reveals the internal relationship between symptoms and their application value. Thus, the work has theoretical and practical implications.

## Methods

### Data Collection

We collected worldwide, long COVID-related, and English-language tweets between May 1, 2020, and December 31, 2021, to create our data set of about 1 million tweets. We used the *Snsrape* module in Python (version 3.8; Python Software Foundation) [16] to scrape the web-based tweet text from tweets that match the keyword "LongCovid." Streaming English-language tweets from multilingual tweets is computationally intensive, because most nonnative English-speaking countries use their native languages rather than English to express their feelings on social media; thus, it requires translating each tweet to English if we wish to analyze them. Therefore, we limited the data set to English-language tweets. The most useful attributes of our data set were ID (Number), time created (DateTime), original text (Text), and language (Text), which we used to filter English-language tweets only. We removed duplicates of tweet text from our data set. This processing necessarily removed any retweets in our data set, which is consistent with our goals of collecting a data set that represents patient experiences. The average tweet length of the initial data set was 32.56 words.

We reduced the data set to 127,848 tweets by limiting the population to those with COVID-19. To do this, we refined the tweets to ensure that all the tweets reflect personal experiences with long COVID. We first considered tweets containing the pronoun "I" and the word "covid" as we wanted to extract tweets from people with COVID-19 or long COVID. Subsequently, we removed tweets containing words that explain users' opinions, as many people discuss long COVID without

necessarily having COVID-19. The set of the words or phrases we considered is listed in [Table 1](#) with the percentage of tweets including the specific word. The phrase “I feel” may include tweets that express the experience of symptoms; however, to remove the possibility of opinion, we removed these tweets. We also observed a number of tweets (3394/148,672, 2.28%) that talked about similar conditions, especially chronic fatigue syndrome. To eliminate the context related to chronic fatigue

syndrome, we removed tweets containing the keyword “cfs.” Our data set was further reduced to identify only the patients discussing their experiences with the symptoms. First, we created a list of long COVID symptoms discussed in different literature sources, as shown in [Table 2](#). We then found the most common symptoms by comparing them with the preprocessed word corpus shown in [Table 3](#).

**Table 1.** The list of words that explain users’ opinions rather than their experience.

Word or phrase	Tweets (N=148,672), n (%)
“opinion”	641 (0.43)
“I believe”	1194 (0.8)
“I think”	6861 (4.61)
“I feel”	2006 (1.35)
“may be” OR “maybe” OR “might”	7582 (5.1)
“perhaps”	750 (0.5)

**Table 2.** List of long COVID symptoms identified by different sources in the literature.

	Symptom	Mayo Clinic	NHS <sup>a</sup>	CDC <sup>b</sup>	WHO <sup>c</sup>	Singh and Reddy [10]
1	Extreme tiredness (fatigue)	✓	✓	✓	✓	✓
2	Shortness of breath or difficulty breathing	✓	✓	✓	✓	✓
3	Cough	✓	✓	✓	✓	✓
4	Joint pain	✓	✓	✓	✓	
5	Chest pain or tightness	✓	✓	✓	✓	✓
6	Problems with memory and concentration (“brain fog”)	✓	✓	✓	✓	✓
7	Difficulty sleeping (insomnia)		✓	✓	✓	✓
8	Muscle pain	✓		✓	✓	✓
9	Headache	✓	✓	✓	✓	✓
10	Fast or pounding heartbeat (heart palpitations) or tachycardia	✓	✓	✓	✓	✓
11	Loss of smell	✓	✓	✓	✓	✓
12	Loss of taste	✓	✓	✓	✓	✓
13	Depression or anxiety					
14	Fever	✓	✓		✓	
15	Dizziness (light-headedness)	✓	✓	✓	✓	✓
16	Worsened symptoms after physical or mental activities	✓		✓		
17	Pins-and-needles feeling		✓	✓	✓	
18	Tinnitus and earaches		✓		✓	✓
19	Diarrhea		✓	✓	✓	
20	Stomach aches		✓	✓		
21	Loss of appetite		✓			✓
22	Sore throat		✓			
23	Rash			✓		
24	Mood changes			✓		
25	Changes in menstrual period cycles			✓	✓	
26	Abdominal pain				✓	✓
27	Neuralgias				✓	
28	Allergies				✓	
29	Body pain					✓
30	Nausea					✓
31	Weakness					✓
32	Numbness					✓

<sup>a</sup>NHS: National Health Service.<sup>b</sup>CDC: Centers for Disease Control and Prevention.<sup>c</sup>WHO: World Health Organization.

**Table 3.** Preprocessed word corpus of stemmed symptoms.

Group	Symptoms
Brain fog	“brain fog,” “brain,” “fog,” “memori,” “mental,” “rememb,” “concentr,” “mind,” “remind,” and “focus”
Fatigue	“fatigu,” “tire,” and “exhaust”
Lung	“lung,” “breathless,” and “breath”
Cannot walk	“cant walk,” “struggl walk,” “unabl walk,” “couldnt walk,” “bare walk,” “unaid walk,” and “stair walk”
Depress	“depress,” “mood,” “stress,” and “anxieti”
Lose weight	“lose weight” and “loss weight”
Insomnia	“cant sleep” and “insomnia”
Diarrhea	“diarrhea” and “diarrhoea”
Dizziness	“dizz” and “lighthead”
Heart	“heart,” “heart palpit,” “tachycardia,” “dysautonomia,” and “arrhythmia”
Others	“headach,” “neck,” “arm,” “muscl pain,” “cough,” “chest pain,” “flu,” “joint pain,” “pain,” “rash,” “fever,” “loss smell,” “loss tast,” “cold,” “earach,” “vomit,” “chill,” “nausea,” “faint,” “gain weight,” “trauma,” “bodi,” “bleed,” “appetit,” “sore throat,” “pin needl,” “numb,” “tinnitu,” “buzz,” “hairfall,” “nose,” “stomach,” “menstrual,” and “abdomin”

### Data Preprocessing

Data preprocessing was mainly used to clean the raw data by following specific steps to achieve better results for further evaluations. We preprocessed our initial data to ensure quality by developing a user-defined preprocessing function based on *Natural Language Toolkit (NLTK)*; a Python library for NLP [15].

The preprocessing plan was as follows. First, we removed the hashtag symbol and its content (eg, COVID-19, @users, and URLs) from the texts because the hashtag symbols and the URLs did not contribute to the text analysis. We also removed all non-English characters (non-American Standard Code for Information Interchange characters) because the study focused on analyzing tweets in English. We then removed repeated words and stop words identified by *NLTK*. Special characters, punctuation, and numbers from the data set were also removed as they did not help the detection of comments with profanity.

We conducted a sentiment analysis to gauge how patients feel about long COVID-related topics, and the results are discussed in the next section. The primary focus of the study was symptom mining; therefore, we created a data set for patients with symptom information using a predefined set of keywords. **Figure 1** shows the word cloud of the set of keywords accounted for initial tweets extraction.

Stemming reduced inflected words to their word stem, base, or root form, whereas tokenization was used to split each sentence into smaller parts of a word. **Figure 2** shows how the data collection, cleaning, and preprocessing steps were conducted and how the final data set with 30,327 tweets was obtained for analysis in the study. After performing cleaning and preprocessing steps, the average tweet length was reduced from 32.56 to 19.07 words. Initially obtained data (original data) and the data considered for the study (filtered data) have been plotted in the same time series plot (**Figure 3**). The filtered data represented the original data appropriately.

**Figure 1.** Word cloud of the list of 73 words after stemming.



Figure 2. Data collection and preprocessing process.

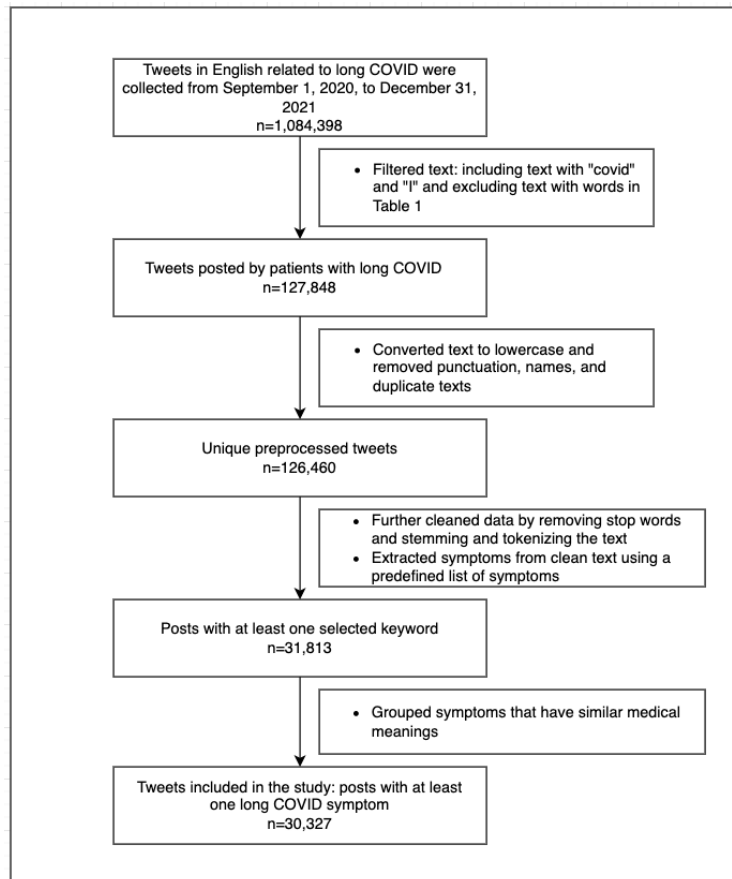
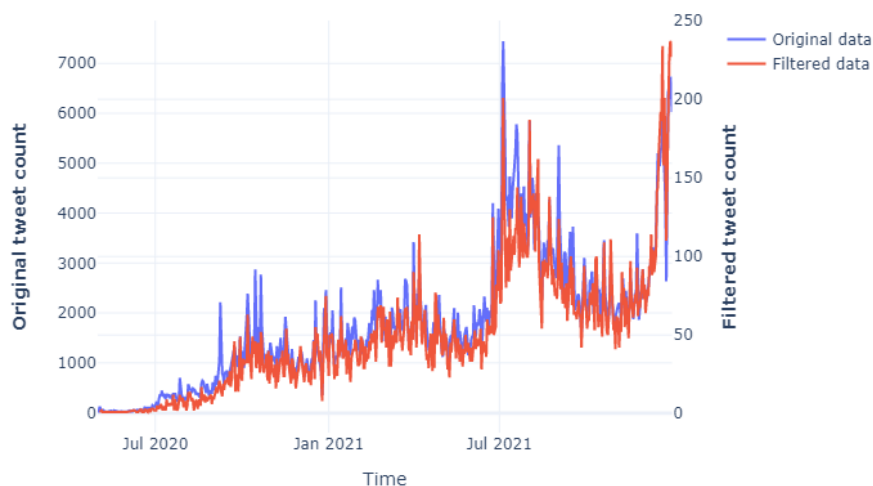


Figure 3. Time series plot for originally obtained data and the data considered for the study.



**Sentiment Analysis**

To measure the sentiment expressed via Twitter on long COVID, we used sentiment analysis, a specific type of NLP, computational linguistics, and text analysis [16,17]. The subjective information from tweets was analyzed and extracted to classify the text into 3 classes, namely positive, negative, and neutral. If the polarity was >0, the text was categorized as

“positive,” and if the polarity was <0, the text was categorized as “negative.” Texts with 0 polarity values were classified as “neutral.” The pie charts in Table 4 show the categorization of the sentiment of each text related to long COVID in 2 stages—for all the posts and posts with at least one long COVID symptoms, respectively. The positive sentiment was reduced from 53.1% (67,153/126,460) to 49.8% (15,102/30,327) when we only considered the tweets with symptoms.

**Table 4.** Classification of the sentiment scores.

	Classes		
	Positive	Negative	Neutral
Sentiment scores of all posts, %	53.1	45.2	1.66
Sentiment scores of posts with at least one long COVID symptom, %	49.8	48.7	1.53

At this stage, we calculated the sentiment polarity of each cleaned and preprocessed tweet using the *TextBlob* library. *TextBlob* is a Python library that supports complex analysis and operations on textual data. It is built on the *NLTK* library [15] and provides a simple API for performing several NLP tasks, such as sentiment analysis, part-of-speech tagging, noun phrase extraction, classification, and translation.

**Collocations**

We also understand that symptoms often appear as more than 1 word in texts. Therefore, finding meaningful symptoms with only 2 words was a particular task in this study. Many valuable text analyses are based on the relationships between words, examining which words tend to follow others immediately or co-occur. Therefore, we analyzed the relationship between 2 words of each bigram in tweets and identified which long COVID symptoms appear as a combination of words. We used the *NLTK* library in Python [15] to identify biwords from the texts and filtered them using the preprepared list of symptoms to only get biwords that can be meaningful symptoms of long COVID when appearing together.

To identify the meaningful biwords, we used the collocation feature in words that reveal a phrase consisting of more than 1 word. Still, these words more commonly co-occur in a given context than their individual word parts. We used several bigram-association measures [18] to filter out the most meaningful collocations.

1. Pointwise mutual information (PMI): the PMI score for 2 words,  $w^1$  and  $w^2$ , is as follows:

$$PMI(w^1, w^2) = \frac{C(w^1, w^2)}{C(w^1)C(w^2)}$$

The main intuition is that it measures how much more likely the words are to co-occur than to occur independently. However,

**Table 5.** Bigram contingency table for a bigram (x y): observed frequencies ( $O_{ij}$ ).  $z^*$  indicates words that are not  $x$  nor  $y$ .  $C(w_i w_j)$  represents the biword count where  $w_i$  and  $w_j$  appear together.  $R_i$  and  $C_j$  are the row and column totals, respectively.  $N$  indicates the total number of biwords in the texts.

	$w^1 = x$	$w^1 \neq x$	
$w^2 = y$	$O_{11} = C(x y)$	$O_{21} = C(z y)$	$R_1$
$w^2 \neq y$	$O_{12} = C(x z)$	$O_{22} = C(z_1 z_2)$	$R_2$
	$C_1$	$C_2$	$N$

the main disadvantage of this method is that it is very sensitive to a rare combination of words. We handled this issue by using the frequency filter of words.

2. Two-tailed  $t$  test with a frequency filter: When conducting a  $t$  test, we can consider testing the below hypothesis with a 5% level of significance.

$$H_0: \text{Words } (w^1, w^2) \text{ occur with probability } \mu$$

versus

$$H_1: \text{Words } (w^1, w^2) \text{ do not occur with probability } \mu$$

Where  $C$  represents the count of each word and  $N$  is the total number of words in the corpus.

The test statistic is as follows:

$$t = \frac{C(w^1, w^2) - \frac{C(w^1)C(w^2)}{N}}{\sqrt{\frac{C(w^1)C(w^2)}{N} \left(1 - \frac{C(w^1, w^2)}{N}\right)}}$$

where,  $\mu$  for Bernoulli trial:

$$\mu = \frac{C(w^1, w^2)}{N}$$

We also avoided the sensitivity to rare cases by filtering the biwords according to the frequencies. Normality assumption is one of the disadvantages of using the  $t$  test.

3. Chi-square test: The null hypothesis of the chi-square test assumes that words ( $w^1, w^2$ ) are independent, just like in the  $t$  test. The chi-square test statistic is computed as follows:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

The observed frequencies ( $O_{ij}$ ) and expected frequencies ( $E_{ij}$ ) can be calculated using the bigram contingency tables presented in [Table 5](#) and [Figure 4](#).

**Figure 4.** Bigram contingency table for a bigram ( $x y$ ): estimated frequencies ( $E_{ij}$ ).  $R_i$  and  $C_j$  are the row and column totals, respectively.  $N$  indicates the total number of biwords in the texts.

	$w^1 = x$	$w^1 \neq x$
$w^2 = y$	$E_{11} = \frac{R_1 C_1}{N}$	$E_{12} = \frac{R_1 C_2}{N}$
$w^2 \neq y$	$E_{21} = \frac{R_2 C_1}{N}$	$E_{22} = \frac{R_2 C_2}{N}$

## Association Rule Mining

In many applications, implications between different situations naturally arise. We refer to these implications as associations. These associations can be discovered and quantified using relational knowledge. “Relational knowledge identifies how concepts/entities are related and how concepts and their relations are defined or described by models” [19]. For example, in the Twitter symptom analysis, some symptoms rules can be determined, such as “2.7% of patients with long COVID with loss of taste also experience loss of smell” or “patients with lung/breathing problems and loss of taste are likely to experience loss of smell with 77% confidence.” These rules are called association rules, and the correlation analysis is known as association mining.

ARM [20] has become an active research field in the data mining community that can solve various problems in health care. Recently, different incremental algorithms have been proposed for mining association rules to discover hidden relationships between symptoms and diseases. In this analysis, ARM was used to obtain insights into the poorly understood LCS by demonstrating the relationships and patterns among the symptoms described in the tweets. We addressed the problem of automatically identifying new and useful symptom patterns in the data of patients with long COVID using an Apriori rule-based data mining algorithm [16]. Recently, other incremental rule-based algorithms, such as Eclat [21], McEclat [22], Direct Hashing and Pruning [23], AprioriTID [24], and MsApriori [25], have also been introduced for mining association rules to discover hidden relationships between item sets and enhance the collection of rare frequent events. We used the Apriori algorithm in this paper as it is a well-understood and well-used algorithm out of this category. Computations and related experiments were done using Python.

We can help define LCS for future research and patient care by describing these relationships among symptoms. These patterns expose the combination of the symptoms that co-occur, as it is helpful to know how 1 symptom or set of symptoms is associated with others. An association rule between a set of symptoms  $X$  and a set of symptoms  $Y$  is expressed in the form  $X \rightarrow Y$ . It is interpreted as “patients with symptoms  $X$  are likely

to have symptoms  $Y$ .” Generally, the effectiveness of discovered rules is measured in terms of support, confidence, and lift.

1. Support: Support indicates how frequently the item set appears in the data set.



2. Confidence: Confidence is the percentage of all transactions satisfying  $X$  that also satisfy  $Y$ .



3. Lift: If the lift is  $>1$ , that lets us know the degree to which 2 occurrences are dependent on one another and makes these rules potentially helpful in predicting the consequences in future data sets.



Thus, based on the analysis of long COVID symptoms, we can mine the association rules among symptoms and quantify their characteristics, such as confidence, support, and lift.

## Results

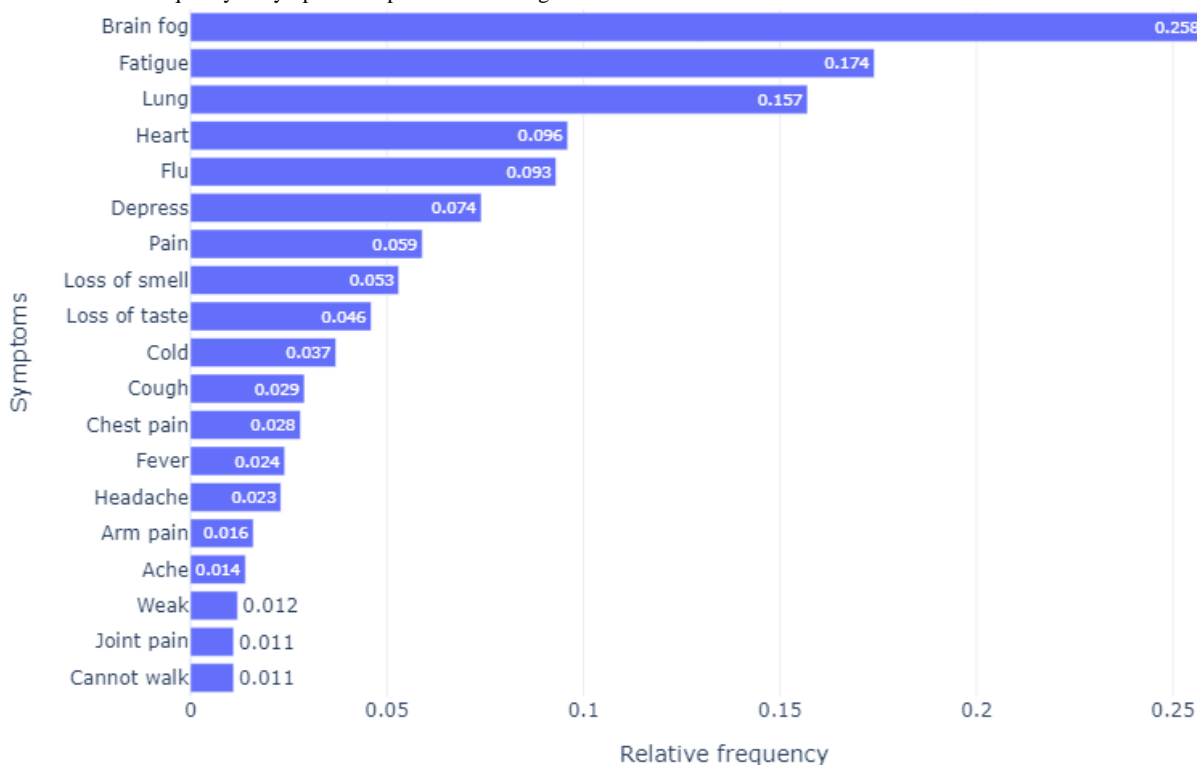
### Symptoms and Medical Conditions Related to Long COVID That Were Discussed on Twitter

Information was extracted for a total of 1,084,398 individual tweets, of which 34,022 had reported long COVID symptoms (Figure 1). The first 5 results obtained from each collocation method discussed in the previous section are shown in Table 6. After observing the results, we determined that the  $t$  test with a filter method achieves acceptable results. We also had to manually select the meaningful biwords with a test statistic  $>5$  and that have a unique meaning only when they are together. We identified 20 such stemmed biwords that appeared together frequently, namely “brain fog,” “loss taste,” “loss smell,” “chest pain,” “cant walk,” “barely walk,” “unable walk,” “struggle walk,” “couldnt walk,” “stair walk,” “joint pain,” “muscle pain,” “leg pain,” “lose weight,” “loss weight,” “gain weight,” “cant sleep,” “sore throat,” “pin & needle,” and “heart palpitation.”

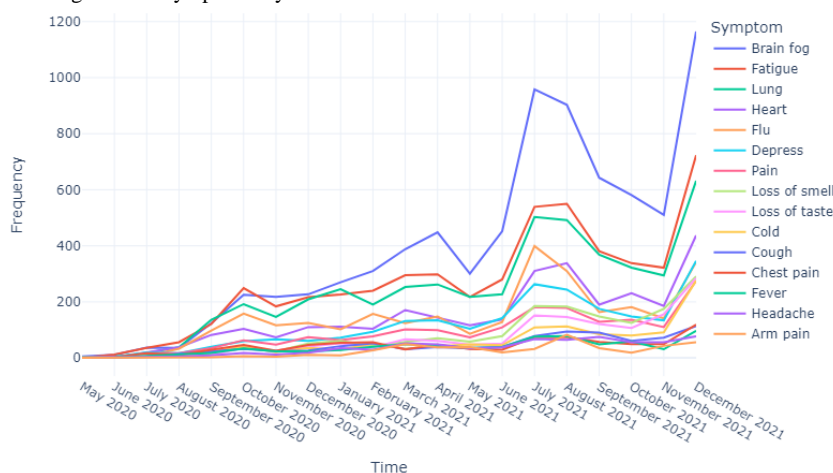




**Figure 6.** Relative frequency of symptoms in patients with long COVID.



**Figure 7.** The 15 most frequent long COVID symptoms by time.



**Symptom Rules**

Our study considered each tweet as a single transaction coming from a single individual. We applied the ARM algorithm to the symptom data considering 1 tweet as 1 transaction and identified symptom rules. The ARM algorithm using symptom transactions aimed to construct frequent item sets, having at least a user-specified threshold. Thus, we set a “confidence” threshold of 0.1 or 10%. We set up a minimum support threshold value above 0.001 and a “lift” greater than 1 for positively correlated rules. We discovered 57 significant rules for the data that included symptom-only information and presented them in Table 7.

The highest confidence level–based detection was used to determine the most significant rules. Among the top 12 rules

that have confidence >0.3, loss of smell and loss of taste were the most common consequent symptoms, followed by lung/breathing problems and fatigue. If a patient had lung/breathing problems and loss of taste, there was a 77% confidence that they had loss of smell. Similarly, patients with fatigue and loss of taste also had loss of smell as a consequent symptom. The top 12 rules are visualized in Figure 8. The following is a description of rule number 11, represented by R11 in a yellow node. There are 3 symptom nodes—“fatigue,” “cough,” and “lung”—represented by green nodes. These 3 nodes form a rule where the antecedents are “fatigue” and “cough,” and the consequent is “lung.” Both the nodes in the antecedent have outgoing links, which are pointed toward the R11 node. Similarly, there is an outgoing link in R11, which is pointing toward the consequent “lung” node.

**Table 7.** List of identified symptom rules.

Rule (R)	Antecedents	Consequents	Support	Confidence	Lift
R0	(loss_taste, lung)	(loss_smell)	0.0028	0.7748	14.5400
R1	(loss_taste, fatigue)	(loss_smell)	0.0028	0.7368	13.8281
R2	(loss_smell, lung)	(loss_taste)	0.0028	0.7107	15.3196
R3	(loss_smell, fatigue)	(loss_taste)	0.0028	0.6614	14.2564
R4	(loss_taste, brain_fog)	(loss_smell)	0.0018	0.5978	11.2192
R5	(loss_taste)	(loss_smell)	0.0272	0.5871	11.0173
R6	(loss_smell)	(loss_taste)	0.0272	0.5111	11.0173
R7	(loss_smell, brain_fog)	(loss_taste)	0.0018	0.4911	10.5847
R8	(heart, pain)	(lung)	0.0012	0.3684	2.3522
R9	(heart, brain_fog)	(lung)	0.0032	0.3542	2.2617
R10	(ache)	(fatigue)	0.0047	0.3471	1.9921
R11	(fatigue, cough)	(lung)	0.0012	0.3083	1.9686
R12	(heart, fatigue)	(lung)	0.0021	0.3014	1.9246
R13	(headache)	(fatigue)	0.0062	0.2675	1.5354
R14	(brain_fog, lung)	(heart)	0.0032	0.2574	2.6915
R15	(heart)	(lung)	0.0223	0.2328	1.4861
R16	(lung, pain)	(fatigue)	0.0016	0.2308	1.3245
R17	(insomnia)	(fatigue)	0.0013	0.2229	1.2791
R18	(muscle_pain)	(fatigue)	0.0019	0.2159	1.2392
R19	(muscle_pain)	(ache)	0.0019	0.2159	15.8929
R20	(cough)	(lung)	0.0061	0.2126	1.3572
R21	(lung, cough)	(fatigue)	0.0012	0.1989	1.1417
R22	(chest_pain)	(lung)	0.0053	0.1891	1.2075
R23	(fever)	(fatigue)	0.0046	0.1871	1.0737
R24	(pain)	(fatigue)	0.0102	0.1736	0.9962
R25	(cold)	(flu)	0.0064	0.1721	1.8483
R26	(lung, pain)	(heart)	0.0012	0.1683	1.7597
R27	(muscle_pain)	(lung)	0.0015	0.1667	1.0641
R28	(weak)	(fatigue)	0.0019	0.1633	0.9374
R29	(trauma)	(brain_fog)	0.0012	0.1586	0.6157
R30	(ache)	(lung)	0.0021	0.1553	0.9918
R31	(fatigue, pain)	(lung)	0.0016	0.1548	0.9886
R32	(ache)	(pain)	0.002	0.1505	2.5553
R33	(heart, lung)	(brain_fog)	0.0032	0.1422	0.5521
R34	(lung)	(heart)	0.0223	0.1421	1.4861
R35	(brain_fog, lung)	(fatigue)	0.0017	0.1421	0.8155
R36	(joint_pain)	(brain_fog)	0.0016	0.142	0.5514
R37	(ache)	(headache)	0.0019	0.1383	6.0025
R38	(ache)	(muscle_pain)	0.0019	0.1383	15.8929
R39	(cough)	(fatigue)	0.004	0.1371	0.7871
R40	(arm_pain)	(fatigue)	0.0021	0.135	0.7749
R41	(muscle_pain)	(pain)	0.0012	0.1326	2.2512



## Discussion

### Principal Findings

The symptoms associated with LCS are still poorly understood. Analyzing social media conversations of patients related to long COVID allows us to understand the frequency and relationship between symptoms. Based on a large amount of Twitter social media data related to LCS, we performed the following 2 tasks in this paper. First, we identified the symptoms and medical conditions related to long COVID that were discussed on the Twitter social media platform. Second, we determined the patterns of symptoms and their associations.

Brain fog, fatigue, and breathing/lung issues were the 3 most common symptoms identified by the analysis. The literature sources verified these reported symptoms [3,4]. With a strong relationship, loss of taste and loss of smell were the most common consequent symptoms. We also identified a substantial number of meaningful symptom rules for a predefined threshold for confidence and support. Only the positively correlated relationships were considered by setting the lift to be greater than 1. The number of strong association rules can be changed according to different minimum thresholds. By finding associations among long COVID symptoms, this work will help physicians identify the behavior of the patients with long COVID. This paper provides new ideas for symptom mining and reveals the relationship between symptoms and their application value. Thus, the work has theoretical and practical importance.

We have used a novel data source, Twitter, and multiple NLP and machine learning techniques to explore the symptoms described by a large undifferentiated population. Different NLP methods, such as sentiment analysis, keyword extraction, and lemmatization, were used to extract information and symptoms from the unstructured text data. Subsequently, we used ARM concepts [20] to expose meaningful relationships between long COVID symptoms. This proof of concept has demonstrated the value of these techniques in describing the frequency of symptoms and the relationships between more common symptoms.

Future research can build on this methodology with clinical data sources such as electronic medical records, adding individual covariates such as sex, age, location, and comorbidities. This research can also be further extended to detect and predict the consequences of a given set of symptoms using word popularity detection methods.

### Limitations

This study is based on web-based Twitter data with limited patient-level variables. No information about the demographics of the tweet authors was available. Furthermore, we only considered patients who shared their experiences with the public in English, as filtering English-language tweets from multilingual tweets is computationally intensive. Another limitation of the study is that results can be affected by misinformation or false conversations on the Twitter platform.

There is a caveat for the confidence metric in the ARM technique when a negative correlation exists between the 2 sets, for instance,  $\sim X \rightarrow Y$ . In most cases, lower support and confidence are preferred when examining negatively correlated rules. The positive symptoms are often clear; however, negative symptoms are subtler and more difficult to recognize and diagnose. As a result, we have not investigated negatively correlated rules. Furthermore, rules discovered by algorithms require clinical validation and verification.

### Conclusion

The most frequent symptoms in our study included brain fog, fatigue, breathing/lung issues, heart issues, flu symptoms, and depression. General pains, loss of smell and taste, cold, cough, chest pain, fever, headache, and arm pain emerged in 1.6% to 5.3% of patients with long COVID. Furthermore, the highest confidence level-based detection with 10% minimum confidence and 0.01% minimum support successfully demonstrates the potential of association analysis and the Apriori algorithm to establish patterns to explore 57 meaningful relationship rules among long COVID symptoms. In this study, to identify the positively correlated symptoms, we only considered the rules with the lift being greater than 1. The results revealed that patients with lung/breathing problems and loss of taste are likely to have loss of smell with 77% confidence.

### Acknowledgments

This study was funded through a Canadian Institutes of Health Research Operating Grant: Emerging COVID-19 Research Gaps and Priorities Funding Opportunity.

### Conflicts of Interest

None declared.

### References

1. Callard F, Perego E. How and why patients made Long COVID. *Soc Sci Med* 2021 Jan;268:113426 [FREE Full text] [doi: [10.1016/j.socscimed.2020.113426](https://doi.org/10.1016/j.socscimed.2020.113426)] [Medline: [33199035](https://pubmed.ncbi.nlm.nih.gov/33199035/)]
2. Yong E. COVID-19 can last for several months. *The Atlantic*. 2020 Jun 04. URL: <https://www.theatlantic.com/health/archive/2020/06/covid-19-coronavirus-longterm-symptoms-months/612679/> [accessed 2022-03-05]
3. Nature Editorial. Long COVID: let patients help define long-lasting COVID symptoms. *Nature* 2020 Oct 07;586(7828):170-170. [doi: [10.1038/d41586-020-02796-2](https://doi.org/10.1038/d41586-020-02796-2)] [Medline: [33029005](https://pubmed.ncbi.nlm.nih.gov/33029005/)]

4. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV, WHO Clinical Case Definition Working Group on Post-COVID-19 Condition. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis* 2022 Apr;22(4):e102-e107 [FREE Full text] [doi: [10.1016/S1473-3099\(21\)00703-9](https://doi.org/10.1016/S1473-3099(21)00703-9)] [Medline: [34951953](https://pubmed.ncbi.nlm.nih.gov/34951953/)]
5. Siddiqui S, Singh T. Social Media its impact with positive and negative aspects. *Int J Comput Appl Tech Res* 2016 Feb 04;5(2):71-75. [doi: [10.7753/ijcatr0502.1006](https://doi.org/10.7753/ijcatr0502.1006)]
6. Fitzgerald R, Radmanesh A. Social media and research visibility. *AJNR Am J Neuroradiol* 2015 Apr;36(4):637 [FREE Full text] [doi: [10.3174/ajnr.A4054](https://doi.org/10.3174/ajnr.A4054)] [Medline: [25034775](https://pubmed.ncbi.nlm.nih.gov/25034775/)]
7. COVID-19: long-term effects. Mayo Clinic. URL: <https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-long-term-effects/art-20490351> [accessed 2022-03-05]
8. Long-term effects of COVID-19 (long COVID). NHS inform. URL: <https://www.nhsinform.scot/long-term-effects-of-covid-19-long-covid> [accessed 2022-03-05]
9. Long COVID or post-COVID conditions. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html> [accessed 2022-03-05]
10. Singh SM, Reddy C. An analysis of self-reported longcovid symptoms on Twitter. medRxiv. Preprint posted online on August 15, 2020 [FREE Full text] [doi: [10.1101/2020.08.14.20175059](https://doi.org/10.1101/2020.08.14.20175059)]
11. Allen J. *Natural Language Understanding*. Redwood City, CA: Benjamin/Cummings Pub. Co; 1995.
12. Chowdhary KR. Natural language processing. In: *Fundamentals of Artificial Intelligence*. New Delhi, India: Springer; Apr 05, 2020:603-649.
13. Chen X, Xie H, Wang FL, Liu Z, Xu J, Hao T. A bibliometric analysis of natural language processing in medical research. *BMC Med Inform Decis Mak* 2018 Mar 22;18(Suppl 1):14 [FREE Full text] [doi: [10.1186/s12911-018-0594-x](https://doi.org/10.1186/s12911-018-0594-x)] [Medline: [29589569](https://pubmed.ncbi.nlm.nih.gov/29589569/)]
14. Stilou S, Bamidis PD, Maglaveras N, Pappas C. Mining association rules from clinical databases: an intelligent diagnostic process in healthcare. In: Patel VL, Rogers R, Haux R, editors. *Studies in Health Technology and Informatics*, vol 84. Amsterdam, the Netherlands: IOS Press; 2001:1339-1403.
15. Bird S, Loper E. NLTK: the natural language toolkit. 2004 Jul Presented at: ACLdemo '04: Proceedings of the ACL 2004 on Interactive poster and demonstration sessions; July 21-26, 2004; Barcelona, Spain p. 31-es. [doi: [10.3115/1219044.1219075](https://doi.org/10.3115/1219044.1219075)]
16. Pang B, Lee L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2008;2(1-2):1-135. [doi: [10.1561/1500000011](https://doi.org/10.1561/1500000011)]
17. Liu B. *Sentiment Analysis and Opinion Mining*. Cham, Switzerland: Springer; May 2012.
18. Andersen G, editor. *Exploring Newspaper Language : Using the Web to Create and Investigate a Large Corpus of Modern Norwegian*. Amsterdam, the Netherlands: John Benjamins; Mar 2012.
19. Halford GS, Wilson WH, Phillips S. Relational knowledge: the foundation of higher cognition. *Trends Cogn Sci* 2010 Nov;14(11):497-505. [doi: [10.1016/j.tics.2010.08.005](https://doi.org/10.1016/j.tics.2010.08.005)] [Medline: [20884275](https://pubmed.ncbi.nlm.nih.gov/20884275/)]
20. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. *SIGMOD Rec* 1993 Jun 01;22(2):207-216 [FREE Full text] [doi: [10.1145/170036.170072](https://doi.org/10.1145/170036.170072)]
21. Zaki MJ, Parthasarathy S, Ogihara M, Li W. Parallel algorithms for discovery of association rules. *Data Min Knowl Discov* 1997;1:343-373. [doi: [10.1023/A:1009773317876](https://doi.org/10.1023/A:1009773317876)]
22. Schlegel B, Karnagel T, Kiefer T, Lehner W. Scalable frequent itemset mining on many-core processors. 2013 Jun Presented at: DaMoN '13: Proceedings of the Ninth International Workshop on Data Management on New Hardware; June 24, 2013; New York, NY p. 1-8. [doi: [10.1145/2485278.2485281](https://doi.org/10.1145/2485278.2485281)]
23. Park J, Yu P, Chen M. Mining association rules with adjustable accuracy. 1997 Jan Presented at: CIKM '97: Proceedings of the sixth international conference on Information and knowledge management; November 10-14, 1997; Las Vegas, NV p. 151--1160. [doi: [10.1145/266714.266886](https://doi.org/10.1145/266714.266886)]
24. Li ZC, He PL, Lei M. A high efficient AprioriTid algorithm for mining association rule. 2005 Nov 07 Presented at: 2005 International Conference on Machine Learning and Cybernetics; August 18-21, 2005; Guangzhou, China p. 1812-1815. [doi: [10.1109/icmlc.2005.1527239](https://doi.org/10.1109/icmlc.2005.1527239)]
25. Lee YC, Hong TY, Lin WY. Mining association rules with multiple minimum supports using maximum constraints. *Int J Approx Reason* 2005 Jul;40(1-2):44-54. [doi: [10.1016/j.ijar.2004.11.006](https://doi.org/10.1016/j.ijar.2004.11.006)]

## Abbreviations

- ARM:** association rule mining
- LCS:** long COVID syndrome
- NLP:** natural language processing
- NLTK:** Natural Language Toolkit
- PMI:** Pointwise mutual information

*Edited by A Mavragani; submitted 14.03.22; peer-reviewed by E Urtnasan, A Pal, H Alhoori; comments to author 09.04.22; revised version received 27.04.22; accepted 18.08.22; published 07.09.22.*

*Please cite as:*

*Matharaarachchi S, Domaratzki M, Katz A, Muthukumarana S*

*Discovering Long COVID Symptom Patterns: Association Rule Mining and Sentiment Analysis in Social Media Tweets*

*JMIR Form Res 2022;6(9):e37984*

*URL: <https://formative.jmir.org/2022/9/e37984>*

*doi: [10.2196/37984](https://doi.org/10.2196/37984)*

*PMID: [36069846](https://pubmed.ncbi.nlm.nih.gov/36069846/)*

©Surani Matharaarachchi, Mike Domaratzki, Alan Katz, Saman Muthukumarana. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 07.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Public Interest and Accessibility of Telehealth in Japan: Retrospective Analysis Using Google Trends and National Surveillance

Takuya Kinoshita<sup>1</sup>, MPH; Takehiro Matsumoto<sup>1,2</sup>, MD, PhD; Naota Taura<sup>2</sup>, MD, PhD; Tetsuya Usui<sup>3</sup>, PhD; Nemu Matsuya<sup>4</sup>, MD, PhD; Mayumi Nishiguchi<sup>1</sup>, NS; Hozumi Horita<sup>1</sup>, NS; Kazuhiko Nakao<sup>1</sup>, MD, PhD

<sup>1</sup>Department of Medical Informatics, Nagasaki University, Nagasaki, Japan

<sup>2</sup>Department of Medical Informatics, Nagasaki University Hospital, Nagasaki, Japan

<sup>3</sup>Department of Laboratory Medicine, Nagasaki University Hospital, Nagasaki, Japan

<sup>4</sup>Department of Neurology, National Hospital Organization Nagasaki Kawatana Medical Center, Kawatana, Japan

**Corresponding Author:**

Takuya Kinoshita, MPH

Department of Medical Informatics

Nagasaki University

1-7-1 Sakamoto

Nagasaki, 852-8501

Japan

Phone: 81 95 819 7529

Email: [takuya\\_kinoshita@nagasaki-u.ac.jp](mailto:takuya_kinoshita@nagasaki-u.ac.jp)

## Abstract

**Background:** Recently, the use of telehealth for patient treatment under the COVID-19 pandemic has gained interest around the world. As a result, many infodemiology and infoveillance studies using web-based sources such as Google Trends were reported, focusing on the first wave of the COVID-19 pandemic. Although public interest in telehealth has increased in many countries during this time, the long-term interest has remained unknown among people living in Japan. Moreover, various mobile telehealth apps have become available for remote areas in the COVID-19 era, but the accessibility of these apps in epidemic versus non-epidemic regions is unknown.

**Objective:** We aimed to investigate the public interest in telehealth during the first pandemic wave and after the wave in the first part of this study, and the accessibility of medical institutions using telehealth in the epidemic and non-epidemic regions, in the second part.

**Methods:** We examined and compared the first wave and after the wave with regards to severe cases, number of deaths, relative search volume (RSV) of telehealth and COVID-19, and the correlation between RSV and COVID-19 cases, using open sources such as Google Trends and the Japanese Ministry of Health, Labour and Welfare (JMHLW) data. The weekly mean and the week-over-week change rates of RSV and COVID-19 cases were used to examine the correlation coefficients. In the second part, the prevalence of COVID-19 cases, severe cases, number of deaths, and the telehealth accessibility rate were compared between epidemic regions and non-epidemic regions, using the JMHLW data. We also examined the regional correlation between telehealth accessibility and the prevalence of COVID-19 cases.

**Results:** Among the 83 weeks with 5 pandemic waves, the overall mean for the RSV of telehealth and COVID-19 was 11.3 (95% CI 8.0-14.6) and 30.7 (95% CI 27.2-34.2), respectively. The proportion of severe cases (26.54% vs 18.16%;  $P < .001$ ), deaths (5.33% vs 0.99%;  $P < .001$ ), RSV of telehealth (mean 33.1, 95% CI 16.2-50.0 vs mean 7.3, 95% CI 6.7-8.0;  $P < .001$ ), and RSV of COVID-19 (mean 52.1, 95% CI 38.3-65.9 vs mean 26.3, 95% CI 24.4-29.2;  $P < .001$ ) was significantly higher in the first wave compared to after the wave. In the correlation analysis, the public interest in telehealth was 0.899 in the first wave and -0.300 overall. In Japan, the accessibility of telehealth using mobile apps was significantly higher in epidemic regions compared to non-epidemic regions in both hospitals (3.8% vs 2.0%;  $P = .004$ ) and general clinics (5.2% vs 3.1%;  $P < .001$ ). In the regional correlation analysis, telehealth accessibility using mobile apps was 0.497 in hospitals and 0.629 in general clinics.

**Conclusions:** Although there was no long-term correlation between the public interest in telehealth and COVID-19, there was a regional correlation between mobile telehealth app accessibility in Japan, especially for general clinics. We also revealed that

epidemic regions had higher mobile telehealth app accessibility. Further studies about the actual use of telehealth and its effect after the COVID-19 pandemic are necessary.

(*JMIR Form Res* 2022;6(9):e36525) doi:[10.2196/36525](https://doi.org/10.2196/36525)

## KEYWORDS

COVID-19; telehealth; telemedicine; public interest; mobile app; correlation; infodemiology, infoveillance; surveillance; Google Trends

## Introduction

Due to the global spread of COVID-19, the government of Japan announced a state of emergency for self-restraint [1]. In the meantime, the World Health Organization recommended telemedicine for treating patients with critical illnesses [2]. Correspondingly, all first visits using telephone or video conferencing were covered temporarily in Japan. Although first visits using telehealth were not covered until then, revisits using telephone had increased after the Tohoku Earthquake in 2011 [3]. Recently, various telehealth apps in mobile devices (or mobile apps) became available in Japan for patients in remote areas. These apps are convenient for patients who have difficulties visiting hospitals, and they attracted the attention of medical practitioners and the public after the emergence of COVID-19 [4]. Along with the demand for telehealth, studies investigating public interest in telehealth increased with the spread of COVID-19 [5,6]. Recently, studies in the fields of infodemiology and infoveillance using web-based sources such as Google Trends are becoming popular for health policy and assessment [7]. Google Trends contain a massive quantity of real-time search interests used for alternative surveillance and health care policy decisions by investigating their correlation [6]. However, most of these studies investigate only the first wave of COVID-19, and therefore, after the first pandemic wave remains unknown [5,6]. There is a study investigating the search term for an alternative surveillance in Japan, using other search engines [8]; however, there is no study focusing on public interest in telehealth in Japan using Google Trends.

In terms of the accessibility of telehealth, many medical institutions started telehealth after COVID-19. In the United States, the accessibility of telehealth increased after the COVID-19 pandemic [9], but the accessibility of medical institutions using telehealth, especially mobile apps, is unknown. The correlation between telehealth accessibility rate and COVID-19 cases may be in line with the hypothesis that epidemic regions have higher accessibility rates. However, we do not know about this correlation across regions. In the first part of this study, we aimed to investigate the long-term correlation between public interest in telehealth and COVID-19 and compared the differences between the first pandemic wave and after the wave. In the second part, we investigated the accessibility of medical institutions using telehealth in the epidemic and non-epidemic regions in Japan.

## Methods

### Data Sources

We used various types of data in this study. For investigating the correlation between public interest in telehealth and COVID-19, we used Google Trends' relative search volume (RSV). RSV is an index of 0-100 representing public interest regarding a specific topic in a certain time and region and with a certain search term. High RSV indicates that many users are searching for a certain term at that point in time in the observed period.

For data about COVID-19, we used nationwide open data from the Japanese Ministry of Health, Labour and Welfare (JMHLW). COVID-19 data include routinely collected daily or cumulative confirmed number of cases, polymerase chain reaction tests, hospitalized cases, severe cases, and deaths from 47 prefectures of Japan; the data were obtained in October 2021. For telehealth accessibility data, a list of telehealth-implemented medical institution names, addresses, telephone numbers, websites, and departments were collected by each prefecture and reported to the government; the data were obtained in October 2020.

To investigate the accessibility of telehealth mobile apps, we used company data from medical institutions listed on the company website. The mobile apps observed were "Clinics," "Curon," "YaDoc," "Carada/Lunaluna," and "Pocket Doctor." These companies had an open list of contract medical institutions in each prefecture on their website. The data were obtained in October 2020. For the denominator used for calculating the proportion of accessible institutions and the prevalence of COVID-19, registered medical institutions and the population in each prefecture from a yearly statistical report called E-STAT were used. The latest annual data about medical institutions were for 2019, and the population data were for 2020, both published in September 2020 and December 2021. The data were obtained in October 2020 and December 2021, respectively.

### Ethical Considerations

The data used in the study were all open data without personal information. Therefore, informed consent was not obtained. The data from JMHLW were also open source. Therefore, no ethical approval was necessary.

### Study Design and Setting

This study had 2 parts. First, we examined the relationship between COVID-19 cases and the RSV of telehealth and COVID-19. The keywords for COVID-19 and telehealth were "corona" and "online-shinryou" in Japanese. These keywords had the largest search volume compared with other potential



search terms (Multimedia Appendix 1). The observed period was 83 weeks (March 1, 2020, to October 2, 2021), with 5 pandemic waves. Both the COVID-19 cases and RSVs were averaged weekly. For comparison, the proportion of severe cases, deaths, and the mean RSV of telehealth and COVID-19 were compared between the first pandemic wave (week 1-13) and after it (week 14-83). The correlation between RSV and COVID-19 cases and the week-over-week change rate were also investigated. Telehealth accessibility between epidemic and non-epidemic regions was then investigated. Tokyo, Kanagawa, Saitama, Chiba, Osaka, Fukuoka, Hokkaido, Ibaraki, Ishikawa, Gifu, Aichi, and Kyoto were selected as epidemic regions (Multimedia Appendix 2). In these prefectures, a state of emergency was declared in the first wave of the COVID-19 pandemic. We included all the institutions from the observed data (Multimedia Appendix 3). Invalid samples such as the ones with missing names, addresses, or domains were excluded. Duplicated data were also excluded from the study. The data were reviewed by 2 researchers and double-checked for compliance. To investigate the regional correlations between COVID-19 cases and telehealth accessibility, we first compared the mean prevalence of COVID-19 cases, severe cases, and deaths; secondly, the mean telehealth accessibility rate of telephone and mobile apps was examined between epidemic and non-epidemic regions. Additionally, for mobile apps, we investigated the regional correlation between the proportion of telehealth-accessible medical institutions and COVID-19 prevalence in each prefecture.

## Statistical Analysis

We assumed the data had no normality. Therefore, the chi-square test for comparing proportions, Mann-Whitney *U* test for comparing means, and Spearman rank-order correlation test for correlation analysis were applied. A *P* value of .05 was considered statistically significant. All analyses were conducted using SPSS software (version 26; IBM Corp).

## Results

### Characteristics and RSV

In the observed 83 weeks with 5 pandemic waves, there were 1,707,581 COVID-19 cases overall: 16,693 in the first pandemic wave and 1,690,888 after the first pandemic wave. The proportion of severe cases (26.54% vs 18.16%; *P*<.001) and deaths (5.33% vs 0.99%; *P*<.001) were significantly higher in the first pandemic wave compared to after the wave (Table 1). The Japanese public interest in telehealth and COVID-19 was the highest in the first wave, and it decreased with time (Figure 1). In terms of increase rate, the highest increase rate was observed in the 5th week (Figure 2). The overall mean RSV of telehealth and COVID-19 was 11.3 (95% CI 8.0-14.6) and 30.7 (95% CI 27.2-34.2), respectively. The mean RSV of telehealth in the first wave was significantly higher compared to the mean RSV of telehealth after the first wave (33.1 vs 7.3; *P*<.001), and this also was the case in the mean RSV of COVID-19 (52.1 vs 26.8; *P*<.001).

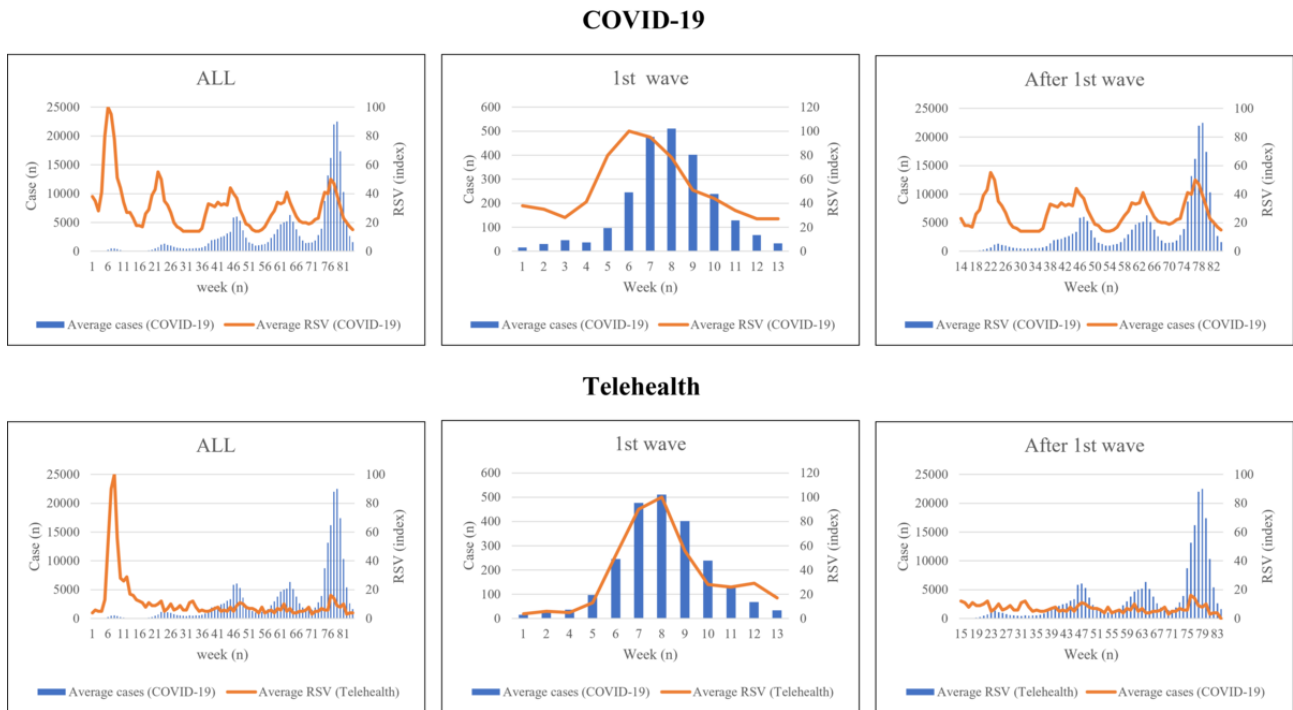
**Table 1.** Basic characteristics of COVID-19 and relative search volume (RSV) in the first wave and after the first wave.

Characteristics	Overall (n=1,707,581)	First wave (n=16,693)	After the first wave (n=1,690,888)	<i>P</i> value
<b>COVID-19 characteristics, n (%)</b>				
Severe cases	311,562 (18.24)	4417 (26.54)	307,145 (18.16)	<.001
Deaths	17,709 (1.03)	891 (5.33)	16,818 (0.99)	<.001
<b>RSV, mean (95% CI)</b>				
Telehealth (search term "online-shinryou")	11.3 (8.0-14.7)	33.1 (16.2-50.0)	7.3 (6.7-8.0)	<.001
COVID-19 (search term "corona")	30.7 (27.2-34.3)	52.1 (38.3-65.9)	26.8 (24.4-29.2)	<.001
<b>Correlation coefficient<sup>a</sup> for telehealth RSV, r</b>				
Mean weekly cases	-0.300	-0.899	-0.208	N/A <sup>b</sup>
Change rate	0.054	0.005	0.137	N/A
<b>Correlation coefficient for COVID-19 RSV, r</b>				
Mean weekly cases	0.152	0.657	0.536	N/A
Change rate	0.428	0.679	0.516	N/A

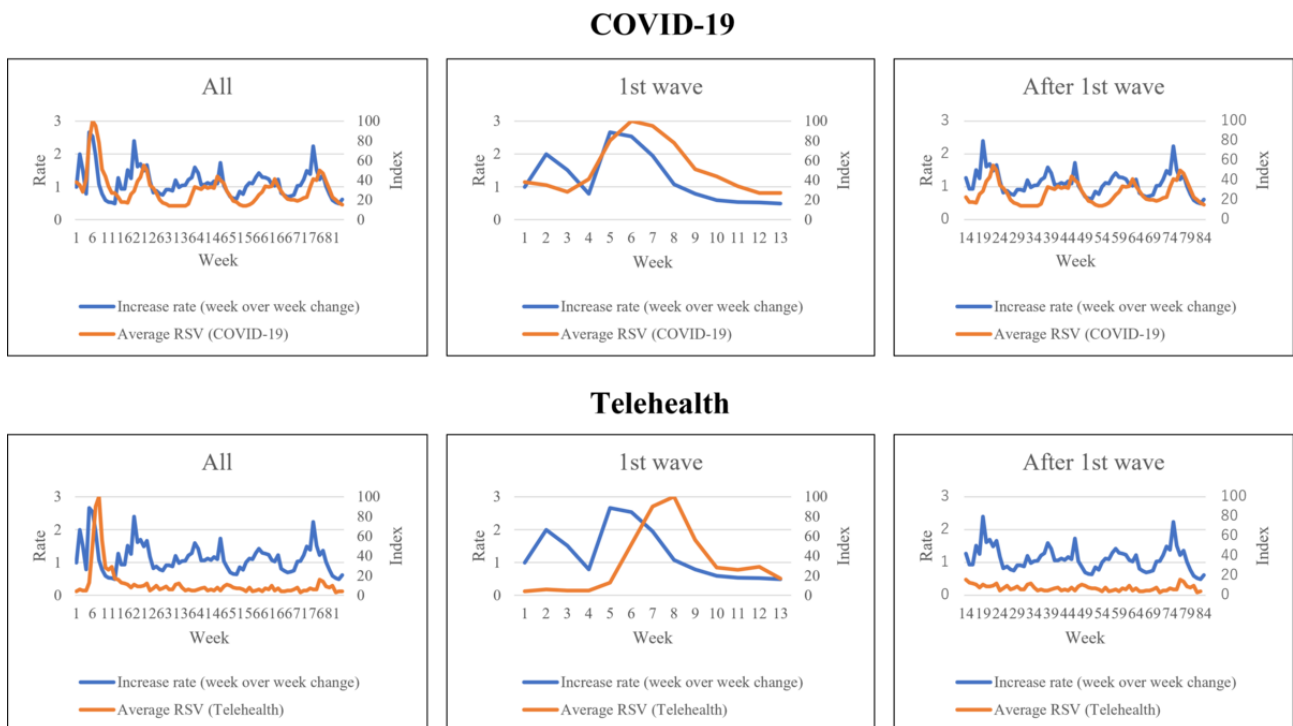
<sup>a</sup>Chi-square test was applied for comparing two proportions and Mann-Whitney *U* test was applied for comparing two means. Spearman rank-order correlation coefficient was applied for public interest of telehealth and COVID-19. *P* value was set at a significant level of .05.

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

**Figure 1.** Public interest and mean weekly COVID-19 cases. RSV: relative search volume.



**Figure 2.** Public interest and week-over-week increase rate of COVID-19 cases. RSV: relative search volume.



**Correlations Between the First COVID-19 Pandemic Wave and After the Wave**

The correlation coefficient between COVID-19 cases and the RSV of COVID-19 was 0.152, but for the RSV of telehealth, it was  $-0.300$  overall (Table 1). When categorizing the period into the first wave and after, the correlation coefficient between the weekly COVID-19 cases and the RSV of telehealth was  $-0.899$ , but that between the cases and the RSV of COVID-19 was 0.657 in the first pandemic wave. After the first wave, the

correlation between the weekly cases of COVID-19 and the RSV of telehealth was  $-0.208$ , but that between the cases and the RSV of COVID-19 was 0.536. When adjusting to week-over-week change rate, the correlation between COVID-19 cases and the RSV of COVID-19 was  $-0.428$ , but that between the cases and the RSV of telehealth was 0.054. Focusing only on the first wave, the correlation between the weekly case of COVID-19 and the RSV of COVID-19 was 0.679, but that between the cases and the RSV of telehealth was 0.005. With regard to change rate, after the first wave, the correlation

between the weekly cases of COVID-19 and the RSV of COVID-19 was 0.516, but that between the cases and the RSV of telehealth was 0.137.

### Accessibility Rate of Telehealth and Regional Correlations

The highest telehealth accessibility was observed in Tokyo with 6.4% in hospitals and 8.8% in general clinics (Figure 3; Multimedia Appendix 4). As shown in Table 2, the accessibility

rate of mobile apps had a significant difference between epidemic and nonepidemic regions in both hospitals (3.8% vs 2%;  $P=.004$ ) and general clinics (5.2% vs 3.1%;  $P<.001$ ), but no significant difference was obtained for all telehealth modes in hospitals (32.8% vs 38.2%;  $P=.24$ ) and general clinics (13.9% vs 14.3%;  $P=.79$ ). In terms of the correlation between mobile app accessibility and the COVID-19 prevalence in each prefecture, the correlation coefficient was 0.497 in hospitals and 0.629 in general clinics (Figure 4).

Figure 3. Accessibility rate of medical institution using mobile telehealth apps (in each prefecture).

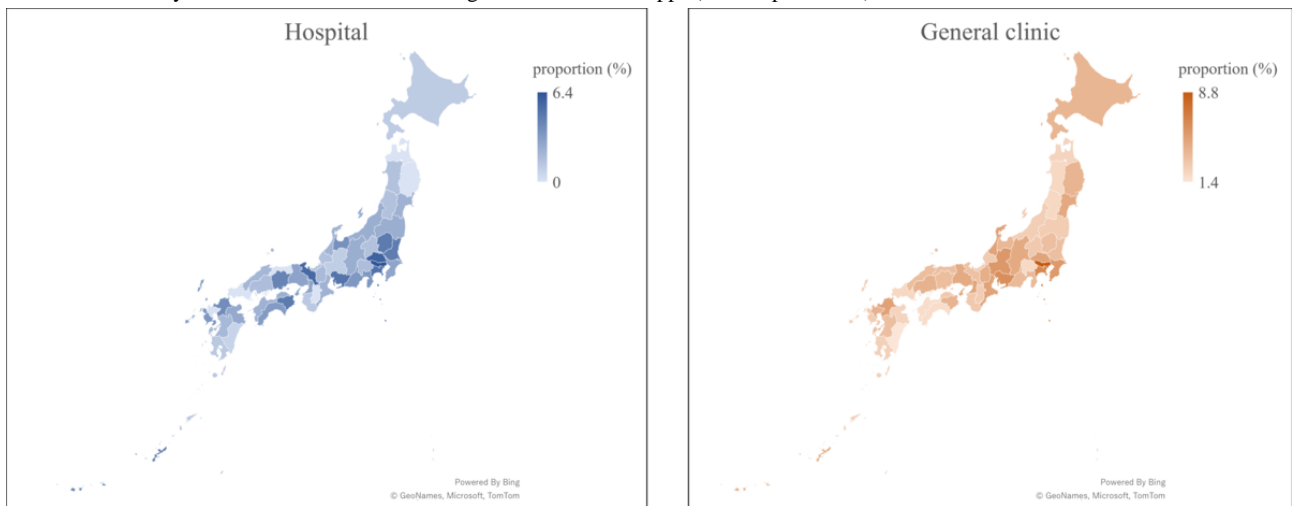
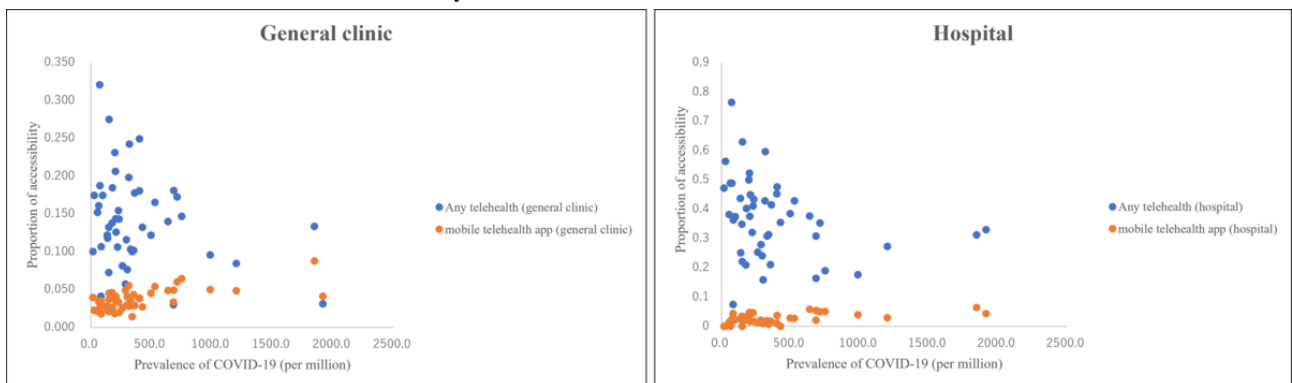


Figure 4. Correlation charts of telehealth accessibility.



**Table 2.** Comparison of basic characteristics of COVID-19 and telehealth accessibility in epidemic and non-epidemic regions.

Characteristics	Total regions (N=47)	Epidemic regions (n=13)	Non-epidemic regions (n=34)	P value
<b>COVID-19/100,000 people (n), mean (95% CI)</b>				
Cases	897.2 (709.6-1084.7)	1495 (1183.1-1806.9)	658.1 (478.6-837.6)	<.001
Severe cases	170.5 (119.0-221.9)	277.3 (160.4-394.2)	127.7 (76.4-179.1)	.006
Deaths	9.1 (7.0-11.1)	16.9 (12.5-21.3)	6 (4.7-7.2)	<.001
<b>Accessibility of all telehealth (%), mean (95% CI)</b>				
Hospital	36.7 (32.8-40.7)	32.8 (26.9-38.6)	38.2 (33.3-43.2)	.24
General clinic	14.2 (12.5-16.0)	13.9 (11.2-16.6)	14.3 (12.1-16.6)	.79
<b>Accessibility of mobile apps (%), mean (95% CI)</b>				
Hospital	2.5 (2.0-3.0)	3.8 (2.7-4.8)	2 (1.5-2.5)	.004
General clinic	3.6 (3.3-4.1)	5.2 (4.3-6.0)	3.1 (2.8-3.4)	<.001
<b>Regional correlation between mobile app accessibility and COVID-19 prevalence, r</b>				
Hospital	0.497 <sup>a</sup>	0.457 <sup>a</sup>	0.496 <sup>a</sup>	
General clinic	0.629 <sup>a</sup>	0.566 <sup>a</sup>	0.627 <sup>a</sup>	

<sup>a</sup>Prefectures declared to be in a state of emergency in the first wave were categorized as epidemic regions (Tokyo, Kanagawa, Saitama, Chiba, Osaka, Hyogo, Fukuoka, Hokkaido, Ibaraki, Ishikawa, Gifu, Aichi, and Kyoto). Student *t* test was applied to compare two independent means for cases of COVID-19, severe cases, and deaths in epidemic and non-epidemic regions. Mann-Whitney *U* Test was applied to compare two independent means for mobile app accessibility rate in hospitals and general clinics. Spearman rank-order correlation was applied for regional correlation of COVID-19 prevalence and the proportion of telehealth availability in each prefecture. *P* value was set at a significant level of .05.

## Discussion

### Principal Findings

In Japan, the public interest in telehealth and COVID-19 cases had a weak negative correlation overall. When adjusting the weekly cases of COVID-19 to week-over-week change rate, the RSV of telehealth did not correlate. Regarding the accessibility of mobile telehealth apps, epidemic regions had significantly higher accessibility rates compared to non-epidemic regions. There was also a positive regional correlation in general clinics. To our knowledge, this study was the first study to investigate the public interest in telehealth, the accessibility of mobile telehealth apps, and the regional correlation between app accessibility and the COVID-19 prevalence in Japan.

### Comparison With Prior Work

In this study, the public interest in telehealth had a weak negative correlation with COVID-19 cases overall, but they were highly correlated in the first wave. This result was similar to previous studies [5,6]. When we adjusted to week-over-week change rate, the correlation coefficient was higher overall for the RSV of COVID-19. Therefore, the public interest may depend on how rapidly the COVID-19 cases increase rather than the weekly cases. Many other possible reasons might have influenced public interest. Considering the Japanese political background, the government announced to reimburse initial visits via telephones and videoconferencing from April 13, 2020. They also reported a nationwide list of telehealth-accessible medical institutions on April 25, 2020. These announcements by the government through the media caused interest among the health care providers and the public. Additionally, a decrease in the proportion of severe cases and deaths after the first wave might have affected public interest.

A study reported that social behavior such as fear could be affected by the uncertainty of the severity and mortality rate [10]. Therefore, with a decrease in the risk of severe illness or death, the restrained patients might have revisited the hospital directly rather than via telehealth. However, the actual use of telehealth or the causation is unclear. Therefore, studies focusing on such aspects need to be conducted in the future.

By looking at previous studies, the accessibility of telehealth increased after the COVID-19 in the United States [11]. In Japan, telehealth accessibility for clinical use was about 1% in small hospitals and general clinics before COVID-19 [3]. This was because treating patients who lived in rural areas was the main purpose of telehealth before COVID-19 [12]. Although the accessibility rate was much higher than previous telehealth use rate in Japan [3], the accessibility rate of mobile apps is still limited (2.5% in hospitals and 3.6% in general clinics). Another study done in Japan showed that the use of telehealth increased among younger people from August to September 2020, compared to April 2020 [13]. Although the study showed an increase in telehealth use, it was a web-based survey, and the observed period was short. Therefore, future studies considering the actual telehealth use after COVID-19 using claims data are necessary.

There was a regional correlation between telehealth accessibility using mobile apps and the COVID-19 cases in Japan, especially for general clinics. This suggests that general clinics using mobile apps for telehealth are more accessible in areas where people must self-restrain under the governmental order. We suspect that general clinics were more likely to apply telehealth to continue medical care under the circumstances. A previous study supports our findings that telemedicine use was higher among those living in urban areas compared to rural areas [13].

Since most of the epidemic regions cover the urban areas of Japan, we believe that similar results were obtained in this study. However, a weak negative regional correlation was observed using all telehealth methods. Since mobile apps are more costly in terms of implementation and management compared to telephones, they may have already been adopted in most medical institutions. In future studies, a comparison of actual telehealth use and the clinical effect of telephones and mobile apps should be investigated.

### Limitations

Several limitations need to be considered in this study. First of all, causation was not investigated in this study. A regression approach should be applied to define the causal relationship. Second, the use of Google Trends does not reflect the entire public interest. However, approximately 70% of all searches are conducted via Google, and it is the most used search engine in Japan [14]. Moreover, the internet use among Japanese people varies by age, and the older population are more likely to have less accessibility to the internet. Therefore, the public interest may not have reflected the actual Japanese population. Furthermore, the keywords chosen in this study may not be appropriate. We believe that the terms “corona” and “online-shinryou” were appropriate based on previous research and the RSV from Google Trend (Multimedia Appendix 1) [8]. Further studies comparing potential keywords are necessary to define the most appropriate search terms. Third, the groups chosen for comparisons such as the periods of the epidemic waves and the prefectures as epidemic regions might be unsuitable. However, we believe that separating the period as first wave and after the first wave was suitable for understanding the differences. Fourth, infectious disease and public interest are said to have lagged in searches among the general population or in reporting confirmed cases for surveillance [15]. However, this study averaged COVID-19 cases by week instead of daily cases. Therefore, the lags have a limited effect on the results.

In addition, there is a time lag for COVID-19 infection, onset, admission, and death [16], and the decision for admission differs by hospital. These aspects were not investigated in this study since we believe that they do not have an immediate effect on public interest. Furthermore, the definition of severe cases may be different between prefectures and JMHLW, since some prefectures did not count patients in the intensive care unit without extracorporeal membrane oxygenation or ventilator. Therefore, the comparison may not be robust. Fifth, we could not investigate the user characteristics of Google Trends or telehealth accessibility. Since the data only show the RSV and the absence of telehealth implementation, further study using real-world data is needed. Finally, the accessibility of mobile telehealth apps through company data does not cover all mobile app use. There are more companies supplying such services that are used by medical institutions. However, we searched the major companies and believe that the results were comprehensive enough to understand the accessibility. Moreover, quite a few medical institutions applied free communication apps such as “Line,” “Zoom,” and “Skype.” These apps are easier to use but have security issues such as protecting personal information [17]. Since those apps were not telehealth apps, we included them in all telehealth groups. We believe that the accessibility rate of mobile apps did not affect the results.

### Conclusions

Although there was no positive long-term correlation between public interest in telehealth and COVID-19, there was a regional correlation between mobile telehealth app accessibility and COVID-19 prevalence in Japan, especially for general clinics. We also revealed that epidemic regions had higher mobile telehealth app accessibility. Further studies about the actual use of telehealth and its effect after the COVID-19 pandemic are necessary.

---

### Acknowledgments

TK analyzed the data and wrote the manuscript. TK and NM reviewed the data. TK, TM, NT, and NM conceptualized and designed the study. TK, TM, NT, NM, TU, MN, HH, and KN revised the manuscript. TK was supported by JSPS KAKENHI grants (21K10331). TM was supported by JSPS KAKENHI grants (20K10322), not related to this study. The content of this paper is solely the responsibility of the authors, and the funders were not involved in the collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. All authors had access to the data in the study and had responsibility for the decision to submit the paper.

---

### Conflicts of Interest

None declared.

---

#### Multimedia Appendix 1

Google Trends Search term (keywords).

[[XLSX File \(Microsoft Excel File\), 14 KB - formative\\_v6i9e36525\\_app1.xlsx](#) ]

---

#### Multimedia Appendix 2

Telehealth accessibility in each prefecture of Japan.

[[DOCX File , 19 KB - formative\\_v6i9e36525\\_app2.docx](#) ]

## Multimedia Appendix 3

Data sample of telehealth accessibility (Tokyo).

[\[PDF File \(Adobe PDF File\), 3558 KB - formative\\_v6i9e36525\\_app3.pdf\]](#)

## Multimedia Appendix 4

Endemic region (prefectures).

[\[PNG File , 120 KB - formative\\_v6i9e36525\\_app4.png\]](#)

## References

1. Velavan TP, Meyer CG. The COVID-19 epidemic. *Trop Med Int Health* 2020 Mar;25(3):278-280 [[FREE Full text](#)] [doi: [10.1111/tmi.13383](https://doi.org/10.1111/tmi.13383)] [Medline: [32052514](https://pubmed.ncbi.nlm.nih.gov/32052514/)]
2. 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](https://doi.org/10.2196/19284)] [Medline: [32501804](https://pubmed.ncbi.nlm.nih.gov/32501804/)]
3. Oh J, Park Y, Jo EC, Kim S. Current status and progress of telemedicine in Korea and other countries. *Healthc Inform Res* 2015 Oct;21(4):239-243 [[FREE Full text](#)] [doi: [10.4258/hir.2015.21.4.239](https://doi.org/10.4258/hir.2015.21.4.239)] [Medline: [26618029](https://pubmed.ncbi.nlm.nih.gov/26618029/)]
4. 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. [doi: [10.2196/24087](https://doi.org/10.2196/24087)] [Medline: [33147166](https://pubmed.ncbi.nlm.nih.gov/33147166/)]
5. Arshad Ali S, Bin Arif T, Maab H, Baloch M, Manazir S, Jawed F, et al. Global interest in telehealth during COVID-19 pandemic: an analysis of Google Trends™. *Cureus* 2020 Sep 16;12(9):e10487 [[FREE Full text](#)] [doi: [10.7759/cureus.10487](https://doi.org/10.7759/cureus.10487)] [Medline: [33083187](https://pubmed.ncbi.nlm.nih.gov/33083187/)]
6. Mavragani A. Tracking COVID-19 in Europe: infodemiology approach. *JMIR Public Health Surveill* 2020 Apr 20;6(2):e18941 [[FREE Full text](#)] [doi: [10.2196/18941](https://doi.org/10.2196/18941)] [Medline: [32250957](https://pubmed.ncbi.nlm.nih.gov/32250957/)]
7. Mavragani A. Infodemiology and infoveillance: scoping review. *J Med Internet Res* 2020 Apr 28;22(4):e16206 [[FREE Full text](#)] [doi: [10.2196/16206](https://doi.org/10.2196/16206)] [Medline: [32310818](https://pubmed.ncbi.nlm.nih.gov/32310818/)]
8. Hisada S, Murayama T, Tsubouchi K, Fujita S, Yada S, Wakamiya S, et al. Surveillance of early stage COVID-19 clusters using search query logs and mobile device-based location information. *Sci Rep* 2020 Oct 29;10(1):18680 [[FREE Full text](#)] [doi: [10.1038/s41598-020-75771-6](https://doi.org/10.1038/s41598-020-75771-6)] [Medline: [33122686](https://pubmed.ncbi.nlm.nih.gov/33122686/)]
9. Hong Y, Lawrence J, Williams D, Mainous I. Population-level interest and telehealth capacity of US hospitals in response to COVID-19: cross-sectional analysis of Google search and national hospital survey data. *JMIR Public Health Surveill* 2020 Apr 07;6(2):e18961 [[FREE Full text](#)] [doi: [10.2196/18961](https://doi.org/10.2196/18961)] [Medline: [32250963](https://pubmed.ncbi.nlm.nih.gov/32250963/)]
10. Wakashima K, Asai K, Kobayashi D, Koiwa K, Kamoshida S, Sakuraba M. The Japanese version of the fear of COVID-19 scale: reliability, validity, and relation to coping behavior. *PLoS One* 2020;15(11):e0241958 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0241958](https://doi.org/10.1371/journal.pone.0241958)] [Medline: [33152038](https://pubmed.ncbi.nlm.nih.gov/33152038/)]
11. Ng BP, Park C. Accessibility of telehealth services during the COVID-19 pandemic: a cross-sectional survey of medicare beneficiaries. *Prev Chronic Dis* 2021 Jul 01;18:E65 [[FREE Full text](#)] [doi: [10.5888/pcd18.210056](https://doi.org/10.5888/pcd18.210056)] [Medline: [34197284](https://pubmed.ncbi.nlm.nih.gov/34197284/)]
12. Miyatake H, Kosaka M, Arita S, Tsunetoshi C, Masunaga H, Kotera Y, et al. Videoconferencing for home care delivery in Japan: observational study. *J Med Internet Res* 2021 Sep 01;23(9):e23539 [[FREE Full text](#)] [doi: [10.2196/23539](https://doi.org/10.2196/23539)] [Medline: [34468333](https://pubmed.ncbi.nlm.nih.gov/34468333/)]
13. Miyawaki A, Tabuchi T, Ong MK, Tsugawa Y. Age and social disparities in the use of telemedicine during the COVID-19 pandemic in Japan: cross-sectional study. *J Med Internet Res* 2021 Jul 23;23(7):e27982 [[FREE Full text](#)] [doi: [10.2196/27982](https://doi.org/10.2196/27982)] [Medline: [34259641](https://pubmed.ncbi.nlm.nih.gov/34259641/)]
14. Search engine market shares in Japan 2021. Statcounter. URL: <https://gs.statcounter.com/search-engine-market-share/all/japan> [accessed 2021-10-15]
15. Shin S, Seo D, An J, Kwak H, Kim S, Gwack J, et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. *Sci Rep* 2016 Sep 06;6:32920 [[FREE Full text](#)] [doi: [10.1038/srep32920](https://doi.org/10.1038/srep32920)] [Medline: [27595921](https://pubmed.ncbi.nlm.nih.gov/27595921/)]
16. García-García D, Vigo MI, Fonfría ES, Herrador Z, Navarro M, Bordehore C. Retrospective methodology to estimate daily infections from deaths (REMEDID) in COVID-19: the Spain case study. *Sci Rep* 2021 May 28;11(1):11274 [[FREE Full text](#)] [doi: [10.1038/s41598-021-90051-7](https://doi.org/10.1038/s41598-021-90051-7)] [Medline: [34050198](https://pubmed.ncbi.nlm.nih.gov/34050198/)]
17. Wakimizu R, Kuroki H, Ohbayashi K, Ohashi H, Yamaoka K, Sonoda A, et al. Perceptions and attitudes toward telemedicine by clinicians and patients in Japan during the COVID-19 pandemic. *Telemed Rep* 2021 Jul 01;2(1):197-204 [[FREE Full text](#)] [doi: [10.1089/tmr.2021.0012](https://doi.org/10.1089/tmr.2021.0012)] [Medline: [35720764](https://pubmed.ncbi.nlm.nih.gov/35720764/)]

## Abbreviations

**JMHLW:** Japanese Ministry of Health, Labour and Welfare**RSV:** relative search volume

*Edited by A Mavragani; submitted 23.01.22; peer-reviewed by S Wei, R Kadel; comments to author 19.04.22; revised version received 24.05.22; accepted 27.06.22; published 14.09.22.*

*Please cite as:*

*Kinoshita T, Matsumoto T, Taura N, Usui T, Matsuya N, Nishiguchi M, Horita H, Nakao K*

*Public Interest and Accessibility of Telehealth in Japan: Retrospective Analysis Using Google Trends and National Surveillance*

*JMIR Form Res 2022;6(9):e36525*

*URL: <https://formative.jmir.org/2022/9/e36525>*

*doi: [10.2196/36525](https://doi.org/10.2196/36525)*

*PMID: [36103221](https://pubmed.ncbi.nlm.nih.gov/36103221/)*

©Takuya Kinoshita, Takehiro Matsumoto, Naota Taura, Tetsuya Usui, Nemu Matsuya, Mayumi Nishiguchi, Hozumi Horita, Kazuhiko Nakao. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 14.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Actions Speak Louder Than Words: Sentiment and Topic Analysis of COVID-19 Vaccination on Twitter and Vaccine Uptake

Murooj Yousef<sup>1\*</sup>, BMultimedia (Hons); Timo Dietrich<sup>1\*</sup>, PhD; Sharyn Rundle-Thiele<sup>1\*</sup>, PhD

Social Marketing @ Griffith, Department of Marketing, Griffith University, Nathan, Australia

\*all authors contributed equally

**Corresponding Author:**

Murooj Yousef, BMultimedia (Hons)

Social Marketing @ Griffith

Department of Marketing

Griffith University

170 Kessels Rd

Nathan, 4111

Australia

Phone: 61 434066942

Email: [m.yousef@griffith.edu.au](mailto:m.yousef@griffith.edu.au)

## Abstract

**Background:** The lack of trust in vaccines is a major contributor to vaccine hesitancy. To overcome vaccine hesitancy for the COVID-19 vaccine, the Australian government launched multiple public health campaigns to encourage vaccine uptake. This sentiment analysis examines the effect of public health campaigns and COVID-19-related events on sentiment and vaccine uptake.

**Objective:** This study aims to examine the relationship between sentiment and COVID-19 vaccine uptake and government actions that impacted public sentiment about the vaccine.

**Methods:** Using machine learning methods, we collected 137,523 publicly available English language tweets published in Australia between February and October 2021 that contained COVID-19 vaccine-related keywords. Machine learning methods were used to extract topics and sentiments relating to COVID-19 vaccination. The relationship between public vaccination sentiment on Twitter and vaccine uptake was examined.

**Results:** The majority of collected tweets expressed negative (n=91,052, 66%) rather than positive (n=21,686, 16%) or neutral (n=24,785, 18%) sentiments. Topics discussed within the study time frame included the role of the government in the vaccination rollout, availability and accessibility of the vaccine, and vaccine efficacy. There was a significant positive correlation between negative sentiment and the number of vaccine doses administered daily ( $r_{267}=.15$ ,  $P<.05$ ), with positive sentiment showing the inverse effect. Public health campaigns, lockdowns, and antivaccination protests were associated with increased negative sentiment, while vaccination mandates had no significant effect on sentiment.

**Conclusions:** The study findings demonstrate that negative sentiment was more prevalent on Twitter during the Australian vaccination rollout but vaccine uptake remained high. Australians expressed anger at the slow rollout and limited availability of the vaccine during the study period. Public health campaigns, lockdowns, and antivaccination rallies increased negative sentiment. In contrast, news of increased vaccine availability for the public and government acquisition of more doses were key government actions that reduced negative sentiment. These findings can be used to inform government communication planning.

(JMIR Form Res 2022;6(9):e37775) doi:[10.2196/37775](https://doi.org/10.2196/37775)

**KEYWORDS**

COVID-19; COVID-19 vaccination; sentiment analysis; public health campaigns; vaccine uptake; Twitter; social media; vaccines



## Introduction

### Background

Vaccination is a widely debated topic, and vaccine hesitancy is a key barrier to national and international immunization efforts; research indicates that certain factors may overcome vaccine hesitancy and influence public sentiment [1]. Vaccine hesitancy is defined as the “delay in acceptance or refusal of vaccination despite the availability of vaccination services” [2]. The lack of trust in vaccines is a major contributor to vaccine hesitancy. Lee et al [3] found that vaccine-hesitant groups are the least likely to express COVID-19 vaccination willingness. Misinformation surrounding vaccine efficacy and side effects can result in the public losing trust in any vaccine, and these negative sentiments can reduce vaccine uptake [4].

Research evidence suggests that vaccine uptake reflects the dynamic interplay between information exchange through interpersonal communications, exposure to public health communication through the media [5], and observations and experiences with government actions (or lack thereof). Scholars have identified social media as an important channel for observing and understanding the public’s thoughts and feelings [6]. People use social media to share their thoughts and beliefs, including views on public health efforts, such as vaccination rollouts, enabling the understanding of public opinions on a large scale.

Public sentiment is defined as people’s opinions, sentiments, evaluations, attitudes, and emotions about a topic [7]. Regarding COVID-19 vaccination, public sentiment could relate to the government vaccination rollout and trust in the vaccine itself. This study investigates key events, such as the launch of public health campaigns, state lockdowns, antivaccination rallies, news related to vaccine efficacy, and vaccine mandates, to identify which government actions can change public vaccination sentiment in Australia.

### Social Media

Social media platforms allow people to express their opinions and emotions about different topics, including health care-related behaviors, such as vaccination. Research suggests that social media is critical in people’s vaccination decision-making process [8]. In terms of willingness to be vaccinated, social media can be a positive or negative influence. Although social media provides governments and health authorities with a platform to disseminate credible public health information that encourages vaccination [9], the same platform can fuel controversial vaccine debates, negatively influencing public opinions and sentiment about vaccines [8]. Previous research has shown that 30%-60% of information about vaccines on social media is antivaccine content [10]. The spread of antivaccination content has serious consequences, including an increase in vaccine hesitancy and delays in vaccine uptake [11,12]. The COVID-19 pandemic presented a contemporary setting to examine vaccine hesitancy, with the proliferation of antivaccination content hindering vaccine acceptance and negatively affecting changes in the vaccine uptake and vaccination sentiment [13,14]. Vaccine hesitancy is not unique to the vaccine for SARS-CoV-2. Vaccine hesitancy is also

evident for other vaccine-preventable diseases, such as measles, mumps, and pertussis, and vaccine hesitancy has been linked to increased disease resurgence. The recent increasing number of deaths related to influenza and viral pneumonia warrants the examination of social media content to determine how it influences vaccine acceptance and uptake [8,15].

Social media is a highly volatile platform where people express positive and negative opinions. Hence, content relating to the COVID-19 vaccine differs daily, reflecting people’s actions and reactions [16]. Different elements may contribute to the change in opinions and sentiment about vaccines, such as national and global events (eg, vaccine trials), legislation (eg, mandating vaccinations), public health campaigns, and vaccine-related news reports [5,8,13]. For example, Tavoschi et al [8] analyzed tweets from 2016 to 2017 to assess the overall sentiments surrounding vaccines in Italy. They found that laws mandating vaccinations for children had a significant negative influence, while disease outbreaks significantly positively influenced public sentiment toward vaccination. Hence, studying the sentiment and opinions of the public during vaccination rollout periods is a good approach for determining vaccination rollout success or failure. Examining public sentiment can also help identify key drivers of positive and negative sentiments toward vaccines, enabling governments and communication agencies to understand which government actions improve the public’s sentiment toward vaccines. This study focuses on the Australian government’s effort to vaccinate the Australian population against COVID-19.

### COVID-19 Vaccination Rollout in Australia

Australia’s strategy to combat COVID-19 was to invest in vaccinating the population against SARS-CoV-2. An effective, safe, accessible, and available vaccine is considered an effective long-term solution to the COVID-19 pandemic. For this study, we defined accessibility as how easily a person could access an appointment for vaccination and availability as the number of available vaccine doses within the country. A critical step in this solution is to vaccinate a high proportion of the population while combating obstacles, such as misinformation, vaccine hesitancy, and lack of trust in government and scientific efforts. To date, only 1 study has analyzed Australian social media content on Twitter from January to October 2020. Kwok et al [17] found that there was general public support for infection control measures (eg, lockdowns) and an overall positive sentiment surrounding the vaccine highlighted by positive emotions, such as trust and anticipation. After that report was published, there were major developments in the vaccination rollout, new communication campaigns were launched, and government announcements were made in Australia. Therefore, additional research considering social media sentiment analysis is warranted.

### Timeline of Key Events During the Australian Vaccination Rollout

The Australian vaccination rollout started in early 2021, with the Australian health minister announcing the federal government’s goal of vaccinating all adult citizens by the end of October 2021 [18]. In May 2021, this goal was reset for the end of 2021 based on the available supply of approved vaccines.

The Therapeutic Goods Administration approved 4 vaccines for Australian use in 2021: the Pfizer-BioNTech vaccine on January 25, the Oxford-AstraZeneca vaccine on February 16, the Janssen vaccine on June 25, and the Moderna vaccine on August 9. The first doses of the COVID-19 vaccine were administered on February 21, 2021, to high-priority groups (ie, front-line workers, health care workers, aged care residents, and workers). This first group received the Pfizer-BioNTech vaccine in highly televised settings to increase trust in the vaccine [19]. On March 22, vaccination of the second-highest-priority groups commenced, focusing vaccination efforts on adults aged >70 years, Aboriginal and Torres Strait Islander people aged >55 years, adults with underlying medical conditions, and emergency service workers. On May 3, adults aged >50 years became eligible for the vaccine, and thousands of participating general practitioners were permitted to administer vaccines in their clinics to expand the scale of the vaccination rollout. As new evidence emerged, suggested changes to the vaccination plan and rollout were observed. For example, the Australian Technical Advisory Group on Immunisation (ATAGI) advised the government to reserve the Pfizer-BioNTech vaccine for people <60 years of age and to administer the more widely available Oxford-AstraZeneca vaccine to adults aged ≥60 years of age. In July, as significant outbreaks of COVID-19 affected Australia's most populated states (New South Wales [NSW] and Victoria), people within communities experiencing such outbreaks were advised to seek vaccination with any available vaccine. At that time, the ATAGI stated, "ATAGI reaffirms our previous advice that in a large outbreak, the benefits of the COVID-19 vaccine AstraZeneca are greater than the risk of rare side effects for all age groups" [20]. Residents aged 16-39 years became eligible to receive the Pfizer-BioNTech vaccine on August 30, and yet more venues were able to administer the vaccines, including vaccination hubs, pharmacies, and community centers. Finally, early adolescents aged 12-15 years became eligible to receive the Pfizer-BioNTech vaccine on September 3. As of October 12, 2021, Australia had administered 31,020,482 doses of COVID-19 vaccines across the country, with 82.8% of people aged 16 years and over having received at least 1 dose and 63.4% having received 2 doses [21]. Although the initial vaccination rollout in Australia was constantly criticized for its slow pace compared to other developed countries, by the end of 2021, vaccination rates in first-dose coverage had surged past many developed nations, including the United States and European Union nations [22]. As of March 2022, 95% of people in Australia over the age of 16 years had received at least 1 dose and over 94% of people had received 2 doses of a COVID-19 vaccine [23].

### Australian Public Health Communication Campaigns

Recent evidence has shown the role of public health campaigns in influencing vaccine-related perceptions and intentions during the COVID-19 pandemic [13]. During the rollout, Australia launched multiple health promotion campaigns aiming to encourage the uptake of vaccines as they became available to the public. The first few campaigns the Australian federal government launched followed a different strategy to the global effort, focusing on informative and data-driven messages rather than emotional and narrative strategies [24]. A few

advertisements were launched informing the public of the availability of the vaccine to certain groups (eg, over 40-year-olds), featuring spokespeople from different health professions, such as Australia's deputy chief medical officer. Such campaigns began in January 2021 and continued to air with multiple versions broadcast until July 2021. With an informational and scientific tone, these campaigns highlighted the safety of the vaccine and encouraged eligible people to register for vaccination. In July, the federal government published its plan to increase the uptake of COVID-19 vaccines through a national COVID-19 campaign, "Operation COVID Shield" [25]. With an AUS \$2 billion (~US \$1.4 billion) budget, the plan featured 3 main work streams: coordinate, motivate, and deliver. Under the motivate work stream, advertising campaign messaging aimed to "motivate eligible people in Australia to receive at least the first dose of the COVID-19 vaccine by 20 December 2021" [25]. Within this plan, 4 public health campaigns were scheduled to run between July and December, 2021. The 4 campaigns are listed in [Multimedia Appendix 1](#).

To assess the effectiveness of these campaigns, the federal government continuously measured sentiment to monitor public confidence in the vaccination rollout [25]. The plan stated, "If...public sentiments are found to be declining, a communications plan will be developed and implemented" [25]. The government assessed sentiments through 2 main data sources, consumer surveys and insights on public sentiments from existing market research programs.

This study aims to assess the change in sentiment and uptake of the COVID-19 vaccine in Australia between February and October 2021. We extended previous research findings into the Australian vaccination rollout, analyzing the effect of public health communication campaigns and actions on vaccine-related sentiment and examining the effect of sentiment on vaccine uptake. Hence, the following research questions (RQs) were examined:

- RQ1. How did vaccination-related sentiment correlate to vaccination rates in Australia between February and October 2021?
- RQ2. How did certain events during the vaccination rollout in Australia relate to changes in the public sentiment toward COVID-19 vaccination?

## Methods

This study utilized machine learning tools to identify public sentiment and analyze emotions related to the vaccination rollout. The analysis of sentiment and vaccine uptake rates was conducted by 1-way ANOVA using SPSS Statistics v. 28 (IBM Corporation).

### Twitter

As 1 of the world's biggest social network platforms, Twitter has 217 million active users, 5.8 million of whom are in Australia [26]. In recent years, Twitter has gained noticeable attention in the scientific literature [6,27]. Twitter users can share instant status updates (tweets) about a range of topics, including health-related content. Twitter is particularly useful

as it shows real-time changes in public opinions, sentiments, and perceptions about vaccinations. Twitter is also an accurate, fast, and cost-effective data source compared to surveys and interviews [6,8,17]. Hence, Twitter was chosen as the data source for this study.

### Data Collection

Multiple methods are available to extract and analyze data from Twitter using artificial intelligence (AI) techniques. Practitioners and scholars are increasingly using social media listening platforms to extract social media data sets and analyze sentiments, topics, and public opinions [27]. Such tools help examine different RQs and inform marketing and communication strategies [28]. An important feature of such tools is the supply of social media data for specific topics, times, and locations. As this study required the collection of tweets and additional data, including date stamps, locations, and links to the tweets collected, the TrackMyHashtag tool was used. The tool provides a dashboard where general analysis of tweet volumes, locations, engagement, and impressions can be viewed and interacted with and the Twitter premium application programming interface (API) service accessed. TrackMyHashtag was used to collect COVID-19 vaccine-related tweets posted between February 1 and October 27, 2021. Following the methodology of Kwok et al [17], retweets, non-English tweets, and tweets with geolocations outside Australia were excluded. Search terms included “vacc OR vax OR vaccine OR vaccination” AND “covid or corona.” Boolean operators were used to ensure that tweets related to “vaccine” and “covid” were collected.

### Sentiment Analysis

Sentiment analysis is defined as “the application of natural language processing, computational linguistics and text analytics to identify and extract subjective information in source materials” [29]. Text analysis involves “information retrieval (IR), lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization and predictive analytics” [28]. An AI-based tool, BytesView, was used through Python to analyze the sentiment of tweets collected for this study [30]. The BytesView tool uses API keys to authenticate requests, enabling access to the tool via multiple applications. The BytesView Python client library was used to integrate the BytesView API into our Python application, enabling the sentiment analysis process. Tweets were rated based on 3 sentiments (positive, negative, and neutral). Further, the R library package *syuzhet* (R Foundation for Statistical Computing) [31], which applies Stanford’s CoreNLP on text against an emotion dictionary, was used to identify specific emotions within tweets (ie, anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) [32]. The valence of a tweet was identified, and then the emotion recognized in the tweet was highlighted. Tweets were scored to reflect the sentiment and emotion within the text.

### Sentiment and Uptake Correlations

Vaccine uptake data are readily available from the Australian Department of Health website. Data from February to October

2021 were exported and listed per day. Pearson correlation coefficient testing (95% CIs) was undertaken to determine relationships between sentiment variables and vaccine uptake. Analysis was undertaken using mean daily, weekly, and monthly values for variables of interest.

### Key Influencing Events

Following the research of Tavošchi et al [8], a set of preselected vaccine-related events (eg, lockdowns) were identified and analyzed to determine their influence on vaccine-related sentiment and uptake. The events that the research team identified included the launch of public health campaigns, lockdowns, COVID-19 vaccine mandates in certain sectors, and changes in travel restrictions. The mean sentiment values from the 5 days preceding (pre-event), during, and after (postevent) each event were analyzed using 1-way ANOVA to assess the variation in sentiment related to these key events. Tukey post hoc testing was undertaken to determine differences within grouping variables across pre-, during-, and postevent time points. All data were analyzed using SPSS Statistics v. 28.0.0.0, and the level of statistical significance was set at  $P < .05$ .

### Topic Modeling

Several Python3 libraries, including Pandas, Regex, Re, and Numpy, were used to develop the latent dirichlet allocation (LDA) topic model. Nltk and spaCy were used for natural language processing, Gensim LDA multicore was used for topic modeling, and PyLDAvis was used for data visualization. The Gensim LDA model has been previously used for topical analysis research [33] and was recently used to analyze COVID-19-related tweets in South Africa [34]. LDA topic modeling was conducted 4 times, first for the whole data set and then for each sentiment group. The model presents 10 topics for each group of tweets and defines the most frequently used words. We evaluated each model based on coherence and perplexity scores. Coherence is the measure of how interpretable topics are for humans, while perplexity measures the efficacy of generative models by measuring the probability of a topic being produced by the model on a data set. A low perplexity score indicates that the model can accurately predict the text corpora of interest [35].

The topic modeling analysis process had multiple steps. First, the data set was loaded into the Pandas data frame via the CSV reader, which extracted the content and sentiment classification from each tweet. Second, we removed all emoji characters and Universal Resource Locators (URLs) from each tweet. Third, stop words were extracted from the nltk library, and additional stop words were added by the research team (see [Multimedia Appendix 2](#)).

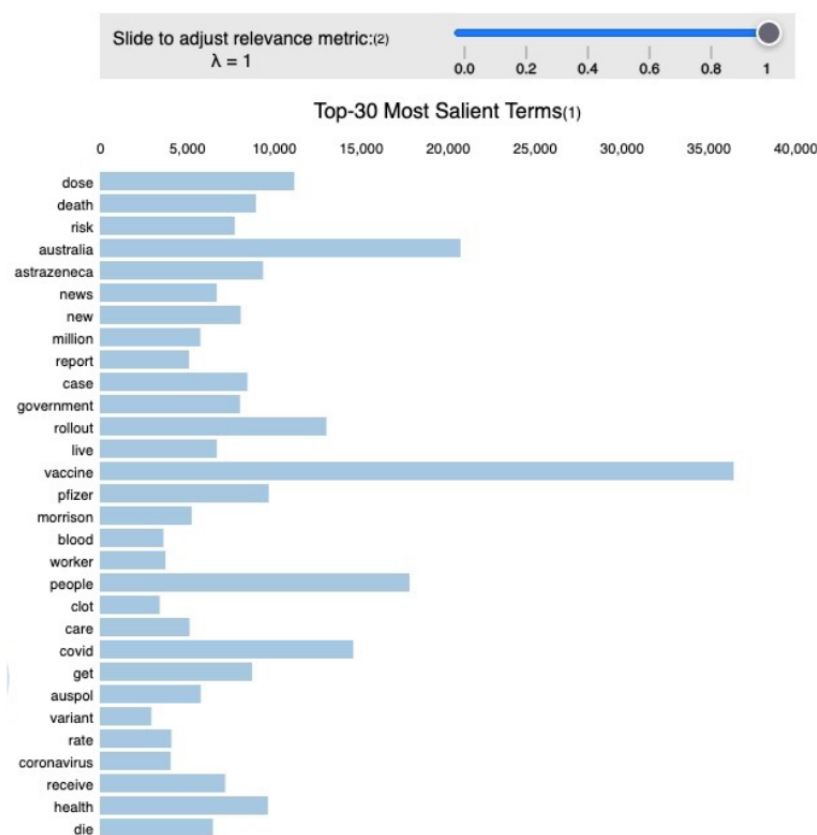
Before running the LDA model, all vocabularies were tokenized with 500 batches for each run; tokens were converted into strings, lemmatized, and then converted back into a string. Subsequently, a word dictionary was created from lemma tokens while filtering extreme values in all tokens. Once a corpus object from lemma tokens was created, the base of the LDA model was initiated using the LdaMulticore algorithm in Python. Model parameters were optimized based on perplexity and coherence scores. Topics were generated, and the final model perplexity



**Table 1.** Topic modeling parameters.

Model parameters	All tweets	Positive	Negative	Neutral
Perplexity	-8.4613	-7.8601	-8.3145	-7.9399
Coherence	0.4859	0.4015	0.4584	0.4069
Number of topics	10	10	10	10
Random state	42	42	42	42
Chunk size	2000	2000	2000	2000
Passes	25	25	25	25
Decay	0.5	0.5	0.5	0.5
Iterations	100	100	100	100
Tweets, n (%)	137,523 (100)	21,686 (16)	91,052 (66)	24,785 (18)

**Figure 3.** The 30 most salient terms in the entire data set of COVID-19 vaccine-related tweets.

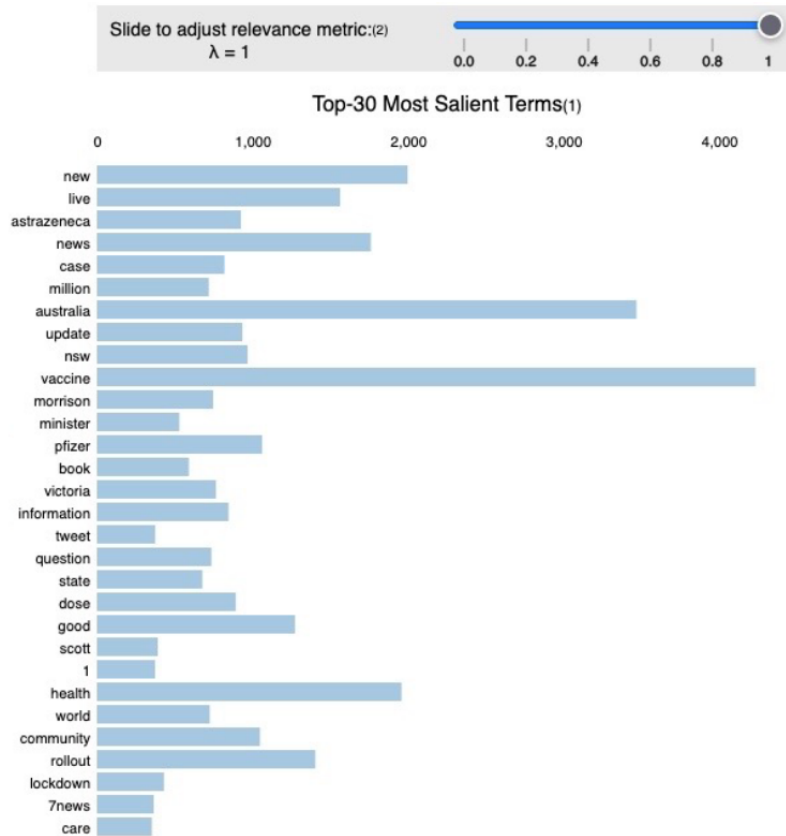


**Positive-Sentiment Tweets**

For this group of tweets, the model achieved a perplexity score of -7.86 and a coherence score of 40% (see Table 1). The list of the extracted topics is shown in Multimedia Appendix 4.

People tweeted positively about being able to book a vaccination appointment, their appreciation for health workers, vaccine safety and related information, vaccine approvals, and vaccination rates. Figure 4 shows the frequency of words used within positive tweets in the data set.

**Figure 4.** The 30 most salient terms in positive-sentiment COVID-19 vaccine-related tweets.

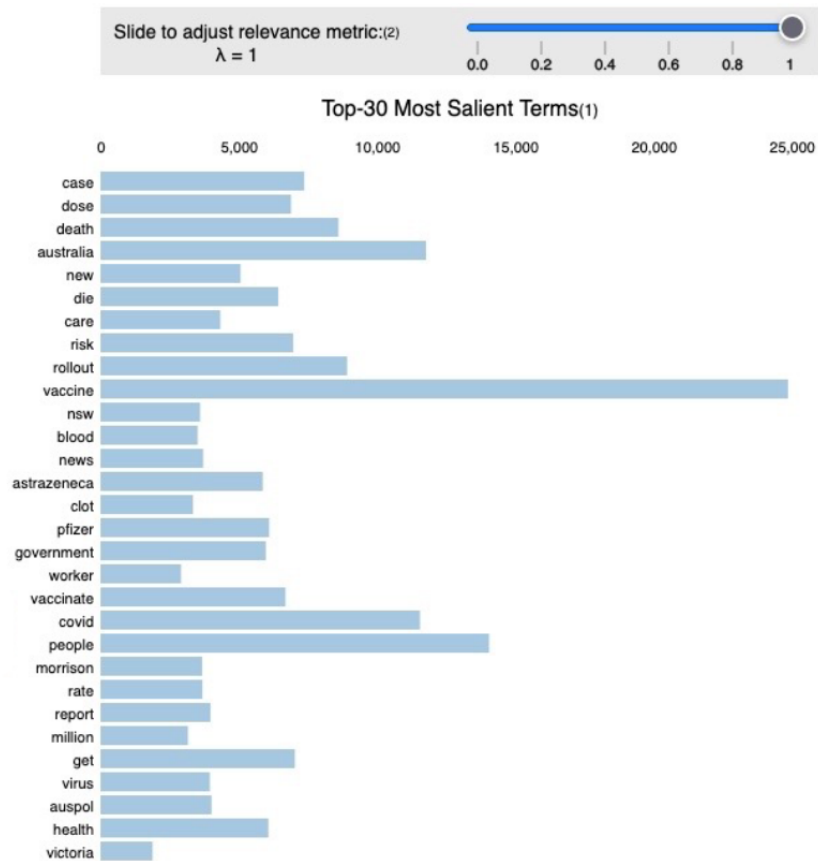


**Negative-Sentiment Tweets**

For all negative tweets within the data set, the model achieved a perplexity score of -8.31 and a coherence score of 45% (see Table 1). Of the 10 identified topics, 5 (50%) were related to the government’s role in the vaccination rollout. These topics

included discussions about vaccine accessibility, availability, approvals, and acquisition. Other topics included border closures, outbreak prevention measures, and vaccine safety and effectiveness. The list of the extracted topics is shown in Multimedia Appendix 5, and the most salient words within this group of tweets are shown in Figure 5.

**Figure 5.** The 30 most salient words within negative-sentiment COVID-19 vaccine-related tweets.

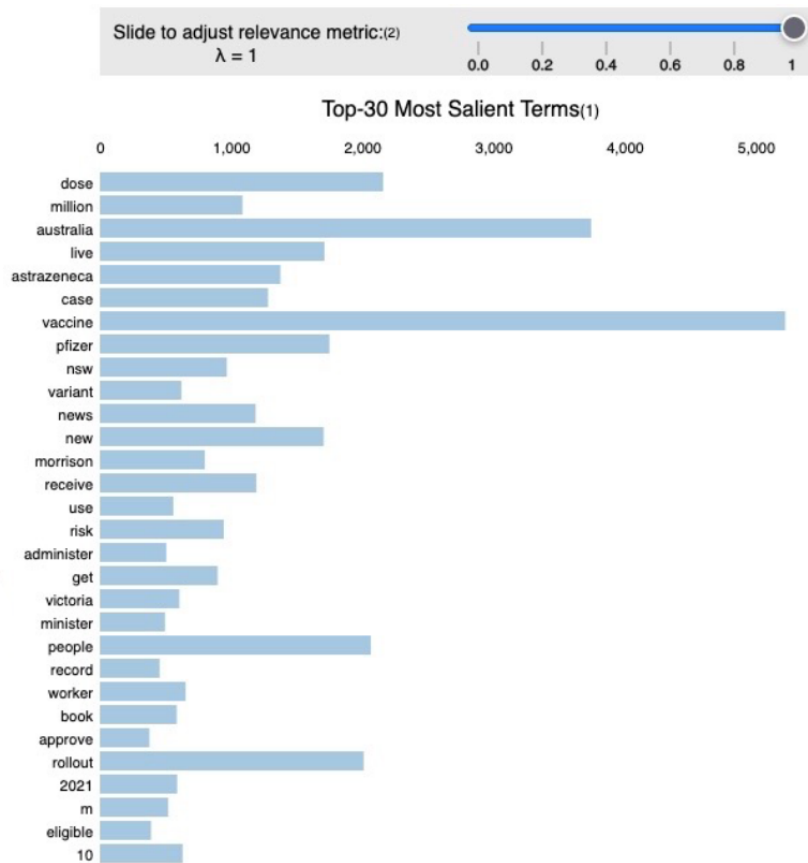


**Neutral-Sentiment Tweets**

For tweets that showed no sentiment and were deemed neutral, the LDA model achieved a perplexity score of  $-7.93$  and a coherence score of 40% (see Table 1). The list of extracted topics (see Multimedia Appendix 6) showed that tweets in this

category discussed vaccine acquisition and accessibility news, vaccine mandates and work requirements, vaccine safety and efficacy, the role of the government in the vaccination rollout, risks of the vaccine, and vaccination approvals for children. The frequency of occurrence of the most salient words is shown in Figure 6.

**Figure 6.** The 30 most salient words within neutral-sentiment COVID-19 vaccine-related tweets.

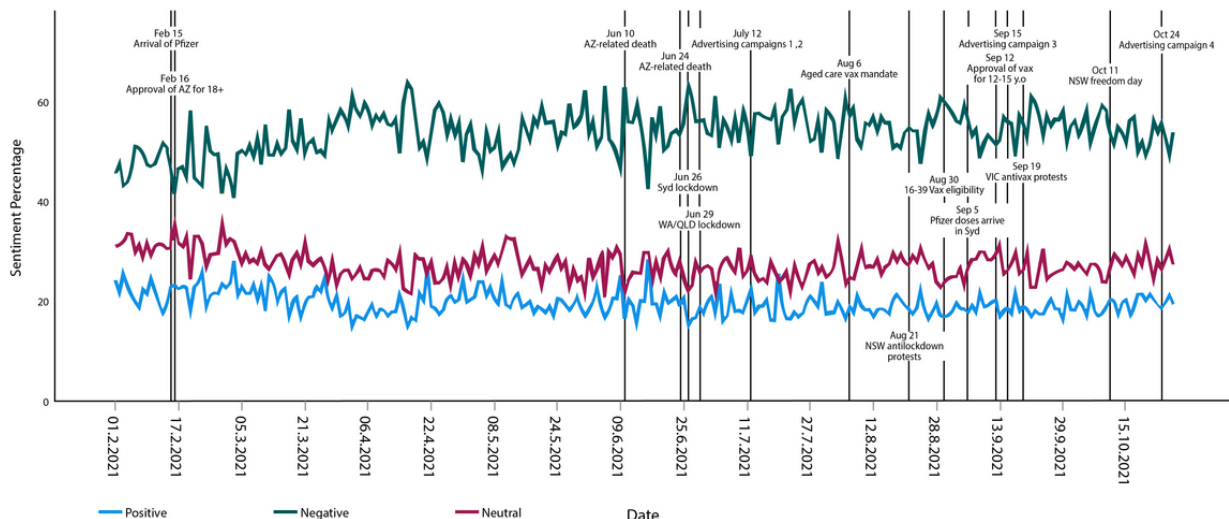


**Sentiment**

Most of the collected tweets expressed negative (n=91,052, 66%) rather than positive (n=21,686, 16%) or neutral (n=24,785, 18%) sentiments. Figure 7 shows the change in sentiment scores of all tweets between February and October 2021, along with the key events in this period. There can be a varying degree of positivity and negativity in each tweet; hence, the higher the sentiment score, the stronger the sentiment expressed in the tweet. Figure 8 shows the positive and negative sentiment levels across the 9 months of the study. Negative sentiment peaked

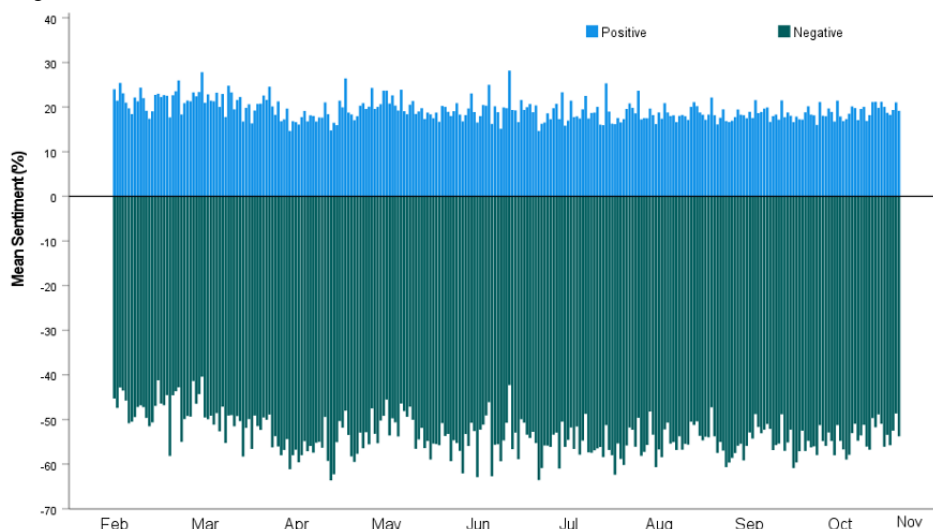
on June 26, 2021, which aligned with the announcement of the greater Sydney lockdown in NSW, and dropped to a minimum on June 16, 2021, which aligned with the availability of the Oxford-AstraZeneca vaccine in selected pharmacies. Positive sentiment peaked on March 3, 2021, when it was announced that 250,000 doses of the Oxford-AstraZeneca vaccine were arriving from Italy. Positive sentiment dropped to its lowest point on April 16, 2021, after the vaccine-related death of a 48-year-old woman, 4 days after receiving the Oxford-AstraZeneca vaccine.

**Figure 7.** Key events and sentiments in COVID-19 vaccine-related tweets over time. NSW: New South Wales; QLD: Queensland; WA: Western Australia; Syd: Sydney; VIC: Victoria; AZ: AstraZeneca; y.o: years old.





**Figure 8.** Positive and negative sentiments in COVID-19 vaccine-related tweets over time.

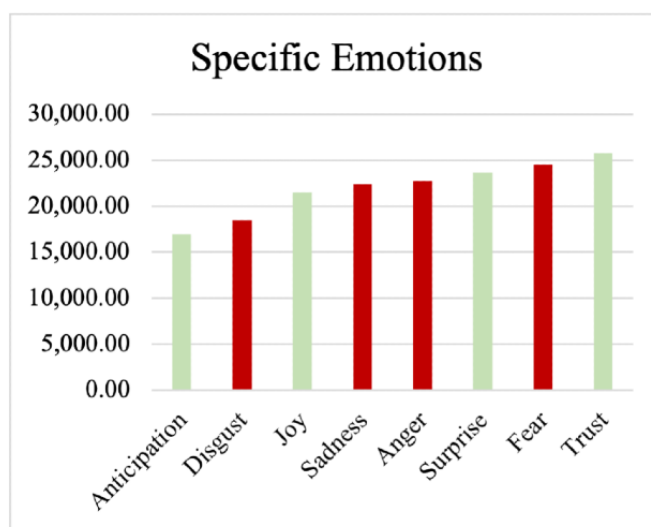


**Emotions**

The emotions of fear, trust, surprise, and anger dominated the tweets collected for this study. The negative emotions expressed included anger (n=23,710, 16%), sadness (n=22,410, 15%), fear (n=21,540, 14%), and disgust (n=18,480, 12%). The most

commonly expressed positive emotions were trust (n=25,730, 17%), anticipation (n=16,950, 11%), joy (n=12,300, 8%), and surprise (n=10,290, 15%). Figure 9 shows the frequency of each emotion in the tweets, with negative emotions in red and positive emotions in green.

**Figure 9.** Emotions detected in COVID-19 vaccine-related tweets.



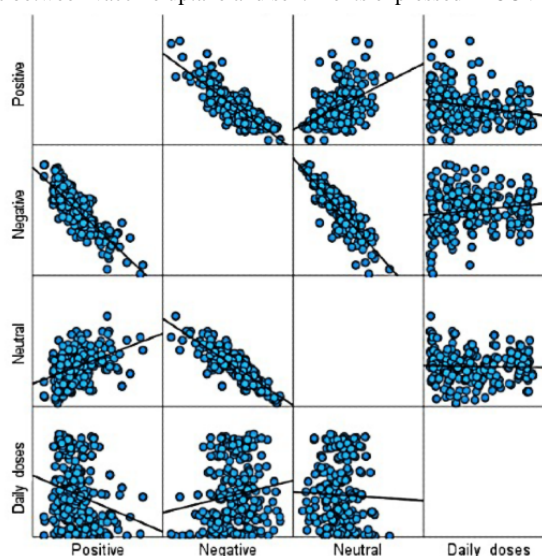
**Uptake and Sentiment Over Time**

When analyzing the relationship between sentiment and COVID-19 vaccine uptake, higher rates of negative sentiment were associated with increased vaccination uptake (see Table 2). Although people may have been expressing negative sentiments about the vaccination rollout, they were still getting vaccinated. A significant positive correlation between negative

sentiment and daily doses administered was evident ( $r_{267}=.15, P=.02$ ). In contrast, when public sentiment was more positive, lower vaccine uptake rates were observed ( $r_{267}=-.23, P<.001$ ). Finally, when sentiment was neutral, there was no significant correlation in vaccine uptake rates ( $r_{267}=-.03, P=.62$ ). Figure 10 shows the correlation results as a scatter plot, identifying a weak correlation between sentiment and uptake.

**Table 2.** Daily sentiment and uptake correlation results.

	Positive	Negative	Neutral	Daily doses
Positive	1.00	N/A <sup>a</sup>	N/A	N/A
Negative	-0.84 <sup>b</sup>	1.00	N/A	N/A
Neutral	0.48 <sup>b</sup>	-0.88 <sup>b</sup>	1.00	N/A
Daily doses	-0.23 <sup>b</sup>	0.15 <sup>c</sup>	-0.03	1.00

<sup>a</sup>N/A: not applicable.<sup>b</sup> $P < .01$ .<sup>c</sup> $P < .05$ .**Figure 10.** Scatter plot of correlation results between vaccine uptake and sentiments expressed in COVID-19 vaccine-related tweets.

## Key Events

The analysis by event was performed on a set of preselected events, including the launch of major national public health campaigns, announcements of vaccine availability, and the start and end of lockdowns. These key events are shown in relation to sentiment change in [Figure 7](#).

## Public Health Campaigns

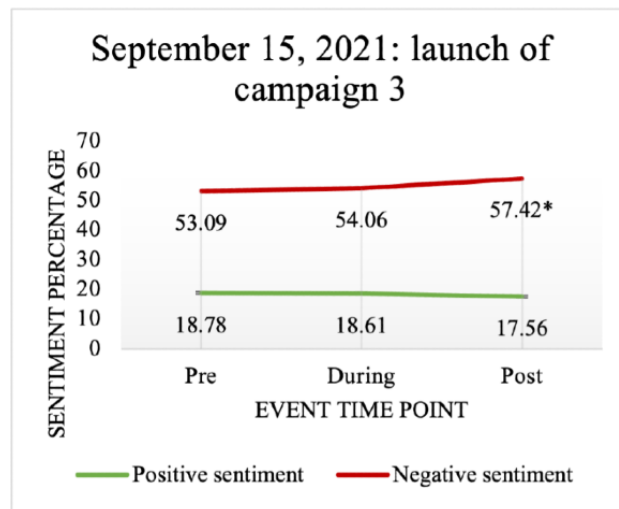
The launch of the 2 first COVID-19 vaccine campaigns on July 12, 2021, had no significant effect on negative or positive sentiments expressed in tweets (see [Table 3](#)). However, [Figure 11](#) shows that the third campaign, “First Things First,” which was launched on September 15, 2021, was associated with a significant increase in negative sentiment postevent compared to pre-event.

**Table 3.** Public health campaigns’ effect on public sentiment.

Event time point	July 12, 2021: campaigns 1 and 2		September 15, 2021: campaign 3	
	Mean (SD)	<i>P</i> value (compared to pre-event)	Mean (SD)	<i>P</i> value (compared to pre-event)
<b>Negative sentiment</b>				
Pre-event	54.26 (32.48)	N/A <sup>a</sup>	53.09 (32.65)	N/A
During the event	54.91 (32.90)	.76	54.06 (32.83)	.52
Postevent	56.12 (32.41)	.11	57.42 (32.39)	<.001
<b>Positive sentiment</b>				
Pre-event	18.77 (22.44)	N/A	18.78 (22.55)	N/A
During the event	19.67 (23.39)	.34	18.61 (22.03)	.95
Postevent	18.53 (22.58)	.93	17.56 (21.39)	.12

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

**Figure 11.** The effect of the September 15, 2021 “First Things First” vaccination campaign on negative and positive sentiments expressed in COVID-19 vaccine-related tweets.

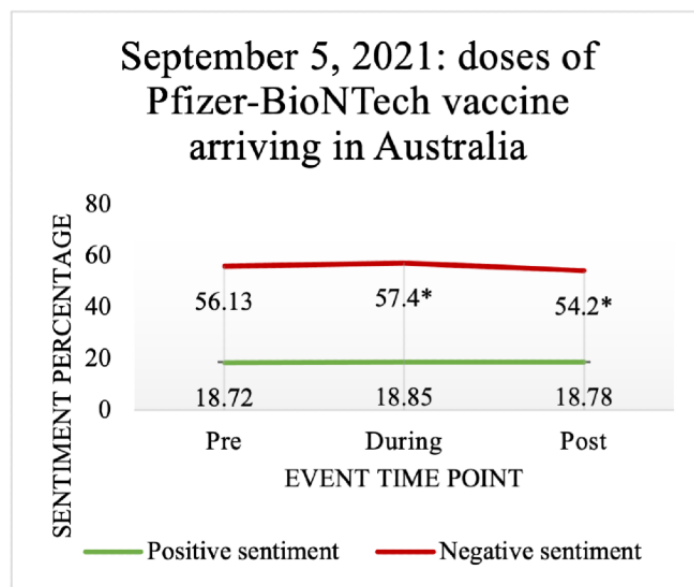


**Vaccine Acquisition– and Approval–Related Events**

The arrival of Pfizer-BioNTech vaccine doses in Australia on February 15, 2021, led to a significant increase in positive sentiment during the event compared to pre-event. The same event led to a significant decrease in negative sentiment during the event compared to pre-event. Similarly, the approval of the Oxford-AstraZeneca vaccine for people aged ≥18 years on February 16, 2021, led to a significant postevent increase in positive sentiment. This same event led to a significant postevent reduction in negative sentiment compared to pre-event. On

September 5, 2021, the arrival of more doses of the Pfizer-BioNTech vaccine in Sydney resulted in a significant reduction in negative sentiment lasting for the 15 days of the event compared to pre-event, as shown in Figure 12. There was a significant decrease in negative sentiment when the Pfizer-BioNTech vaccine was approved on August 30, 2021, for people aged 16-39 years. However, when vaccination was approved on September 12, 2021, for people aged 12-15 years, there was a significant increase in negative sentiment compared to pre-event (see Tables 4-8).

**Figure 12.** The effect of vaccine availability on negative and positive sentiments expressed in COVID-19 vaccine-related tweets.



**Table 4.** Vaccine acquisition and approval effect on public sentiment (February 15, 2021: arrival of Pfizer-BioNTech doses in Australia).

Event time point	Mean (SD)	<i>P</i> value (compared to pre-event)
<b>Negative sentiment</b>		
Pre-event	49.19 (32.77)	N/A <sup>a</sup>
During the event	45.04 (32.46)	<.001
Postevent	47.99 (32.98)	.36
<b>Positive sentiment</b>		
Pre-event	20.38 (23.68)	N/A
During the event	22.69 (26.02)	.003
Postevent	21.86 (25.06)	.07

<sup>a</sup>N/A: not applicable.**Table 5.** Vaccine acquisition and approval effect on public sentiment (February 16, 2021: approval of Oxford-AstraZeneca for people aged ≥18 years).

Event time point	Mean (SD)	<i>P</i> value (compared to pre-event)
<b>Negative sentiment</b>		
Pre-event	48.93 (32.77)	N/A <sup>a</sup>
During the event	47.28 (32.63)	.16
Postevent	46.86 (33.06)	.04
<b>Positive sentiment</b>		
Pre-event	20.38 (23.88)	N/A
During the event	21.65 (25.03)	.15
Postevent	22.29 (25.54)	.01

<sup>a</sup>N/A: not applicable.**Table 6.** Vaccine acquisition and approval effect on public sentiment (September 5, 2021: doses of Pfizer-BioNTech vaccine arriving in Sydney from the United Kingdom).

Event time point	Mean (SD)	<i>P</i> value (compared to pre-event)
<b>Negative sentiment</b>		
Pre-event	56.13 (32.46)	N/A <sup>a</sup>
During the event	57.40 (31.88)	<.001
Postevent	54.20 (31.94)	.001
<b>Positive sentiment</b>		
Pre-event	18.72 (22.36)	N/A
During the event	18.85 (22.45)	.97
Postevent	18.78 (22.55)	.99

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

**Table 7.** Vaccine acquisition and approval effect on public sentiment (August 30, 2021: Pfizer-BioNTech vaccine approved for people aged 16-39 years).

Event time point	Mean (SD)	P value (compared to pre-event)
<b>Negative sentiment</b>		
Pre-event	56.77 (32.19)	N/A <sup>a</sup>
During the event	57.40 (31.88)	.76
Postevent	54.20 (31.94)	.01
<b>Positive sentiment</b>		
Pre-event	17.65 (21.26)	N/A
During the event	17.83 (21.33)	.95
Postevent	18.75 (22.15)	.17

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

**Table 8.** Vaccine acquisition and approval effect on public sentiment (September 12, 2021: vaccinations approved for people aged 12-15 years).

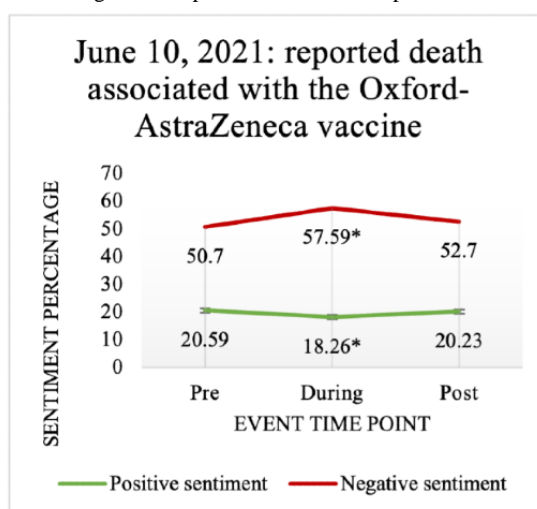
Event time point	Mean (SD)	P value (compared to pre-event)
<b>Negative sentiment</b>		
Pre-event	52.06 (32.11)	N/A <sup>a</sup>
During the event	54.34 (32.48)	.04
Postevent	54.82 (32.97)	.01
<b>Positive sentiment</b>		
Pre-event	19.26 (22.87)	N/A
During the event	18.11 (21.79)	.16
Postevent	18.55 (21.94)	.48

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

### Vaccine Side Effects and Death Events

On June 10, 2021, a death was attributed to the Oxford-AstraZeneca vaccine, resulting in a significant increase in negative sentiment during the event compared to pre-event, as shown in Figure 13. Similarly, positive sentiment significantly

decreased during the event compared to pre-event. No significant effect was found on postevent sentiment. On June 24, 2021, another death was associated with the administration of the Oxford-AstraZeneca vaccine, resulting in similar trends in increased negativity and decreased positivity during the event compared to pre-event (see Table 9).

**Figure 13.** The effect of a vaccine-related death on negative and positive sentiments expressed in COVID-19 vaccine-related tweets.

**Table 9.** Vaccine side effects’ and deaths’ influence on public sentiment.

Event time point	June 10, 2021: Oxford-AstraZeneca vaccine-related death		June 24, 2021: Oxford-AstraZeneca vaccine-related death	
	Mean (SD)	P value (compared to pre-event)	Mean (SD)	P value (compared to pre-event)
<b>Negative sentiment</b>				
Pre-event	50.70 (33.36)	N/A <sup>a</sup>	53.21 (57.91)	N/A
During the event	57.59 (32.35)	<.001	57.91 (32.04)	<.001
Postevent	52.70 (33.71)	.10	55.90 (32.04)	.12
<b>Positive sentiment</b>				
Pre-event	20.59 (24.65)	N/A	19.60 (23.31)	N/A
During the event	18.26 (22.54)	.003	16.86 (20.06)	<.001
Postevent	20.23 (23.72)	.86	18.73 (22.18)	.34

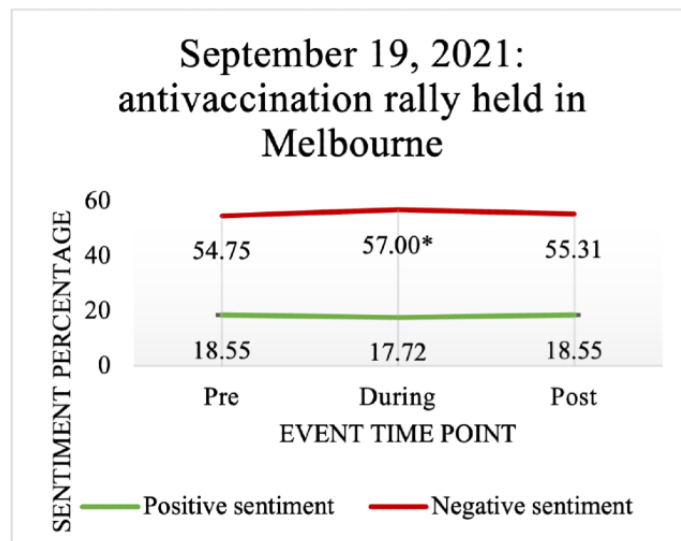
<sup>a</sup>N/A: not applicable.

**Antivaccination and Antilockdown Rallies**

Multiple antivaccination rallies were held across Australia in 2021. On September 19, 2021, a violent antivaccination rally occurred in Melbourne, Victoria—a city that had experienced the longest lockdown periods in the world—associated with a significant increase in negative sentiment during the event compared to pre-event, as shown in Figure 14. However, no

significant effect on sentiment was found when comparing pre- and postevent negative sentiment or in comparisons of positive sentiment. Similarly, in NSW, when antilockdown and antivaccination rallies happened on August 21, 2021, negative sentiment significantly increased from pre- to postevent. Additionally, positive sentiment significantly decreased from pre- to postevent. No significant effect on either sentiment was found during the event (see Table 10).

**Figure 14.** The effect of an antivaccination rally on negative and positive sentiments expressed in COVID-19 vaccine-related tweets.



**Table 10.** Antivaccination and antilockdown rallies' effect on public sentiment.

Event time point	September 19, 2021: Melbourne antivaccination rally		August 21, 2021: NSW <sup>a</sup> antilockdown and antivaccination rallies	
	Mean (SD)	<i>P</i> value (compared to pre-event)	Mean (SD)	<i>P</i> value (compared to pre-event)
<b>Negative sentiment</b>				
Pre-event	54.75 (32.65)	N/A <sup>b</sup>	18.55 (22.20)	N/A
During the event	57.00 (32.55)	.04	17.72 (21.41)	.38
Postevent	55.31 (32.67)	.82	18.55 (22.14)	.99
<b>Positive sentiment</b>				
Pre-event	52.27 (32.50)	N/A	19.46 (23.13)	N/A
During the event	52.73 (32.58)	.87	18.80 (22.38)	.55
Postevent	57.95 (31.72)	<.001	17.35 (20.62)	.002

<sup>a</sup>NSW: New South Wales.

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

### Lockdowns and Legislations

Lockdowns across different states had varying effects on sentiment. When a snap lockdown occurred in Queensland (QLD) and Western Australia (WA) on June 29, 2021, postevent negative sentiment significantly decreased compared to pre-event, indicating public support for the actions taken by these state governments. No significant effect on sentiments during the event was found. Positive sentiment also significantly increased throughout the events. In contrast, when a lockdown was announced for Greater Sydney on June 26, 2021, negative sentiment significantly increased during the 15 days of the event

and postevent compared to pre-event, with the highest level of negative sentiment observed during the event. Positive sentiment significantly decreased during this lockdown period compared to pre-event. No significant effect on positive sentiment was found postevent compared to pre-event. When the lockdown was lifted in NSW on October 11, 2021, negative sentiment significantly decreased, but there was no significant effect on positive sentiment compared to pre-event (see [Table 11](#)).

No significant changes in negative or positive sentiment were observed when COVID-19 vaccination was mandated for aged care workers on August 6, 2021 (see [Table 12](#) and [Figure 15](#)).

**Table 11.** Lockdowns' effect on public sentiment.

Event time point	June 29, 2021: QLD <sup>a</sup> and WA <sup>b</sup> lockdowns		June 26, 2021: Greater Sydney lockdown		October 11, 2021: Sydney lockdown lifted	
	Mean (SD)	<i>P</i> value (compared to pre-event)	Mean (SD)	<i>P</i> value (compared to pre-event)	Mean (SD)	<i>P</i> value (compared to pre-event)
<b>Negative sentiment</b>						
Pre-event	57.91 (32.04)	N/A <sup>c</sup>	52.67 (32.87)	N/A	55.90 (32.07)	N/A
During the event	55.90 (32.80)	.09	58.43 (31.63)	<.001	52.67 (32.74)	.001
Postevent	53.94 (32.64)	<.001	54.87 (33.12)	.05	52.42 (33.09)	<.001
<b>Positive sentiment</b>						
Pre-event	16.86 (20.06)	N/A	19.70 (23.37)	N/A	18.41 (22.02)	N/A
During the event	18.73 (22.18)	.01	16.67 (19.86)	<.001	19.34 (23.01)	.32
Postevent	19.02 (32.64)	.002	19.34 (23.05)	.83	19.45 (23.28)	.23

<sup>a</sup>QLD: Queensland.

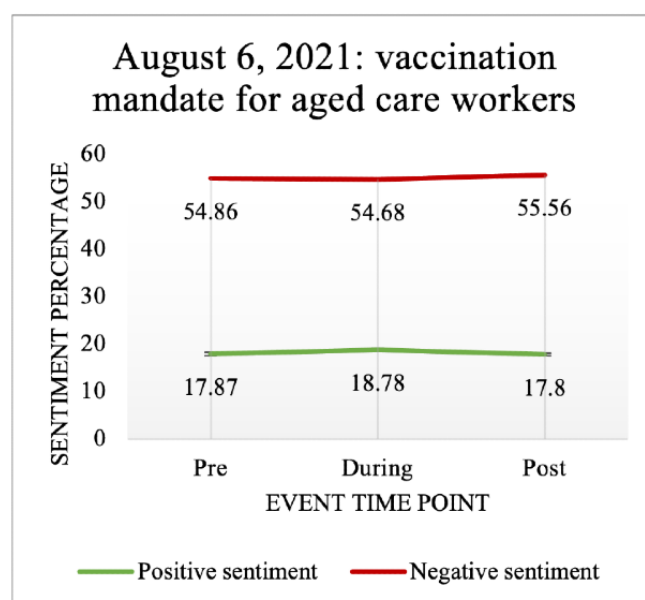
<sup>b</sup>WA: Western Australia.

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

**Table 12.** Vaccination mandate's (August 6, 2021: for aged care workers) effect on public sentiment.

Event time point	Mean (SD)	P value (compared to pre-event)
<b>Negative sentiment</b>		
Pre-event	54.86 (32.60)	N/A <sup>a</sup>
During the event	54.68 (32.60)	.98
Postevent	55.56 (31.85)	.73
<b>Positive sentiment</b>		
Pre-event	17.87 (21.79)	N/A
During the event	18.78 (22.32)	.31
Postevent	17.80 (21.39)	.99

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

**Figure 15.** The effect of COVID-19 vaccination mandates for aged care workers on negative and positive sentiments expressed in COVID-19 vaccine-related tweets.

## Discussion

### Principal Findings

This study aimed to examine the relationship between public sentiment and COVID-19 vaccine uptake and between public sentiment and key actions and events. Key findings of this study included a negative correlation between sentiment and vaccine uptake and a significant change in sentiment observed for vaccine accessibility and availability events. Key topics within the analyzed data set showed that the population was discontented with the government's effort in managing the vaccination rollout. This study contributes to our understanding of the role of public sentiment in vaccine uptake and overcomes the limitations of conventional survey methods [36]. Although big data and semantic analysis have been used for COVID-19 research, such analysis is mostly used to understand public discussions about the pandemic and related control measures [17] and public behavior, such as panic buying [37]. There remains a research gap in using machine learning methods to understand COVID-19 vaccination sentiment and actions beyond the response bias that typically occurs in a representative sample

population [38]. Further, this study evaluated the effects of different Australian public health campaigns on public sentiment, improving our understanding of these campaigns' effectiveness and providing recommendations for future campaign improvements. Similarly, implications of government efforts during a vaccination rollout were discussed, highlighting the need to align messages with the accessibility and availability of vaccines.

### Sentiment as an Indicator of Vaccine Uptake

Previously, research relied on self-reported measures to understand vaccine sentiment and the willingness to be vaccinated [36]. This study extended on previous research that applied machine learning methods to illuminate the limitations of self-reported measures. Although the literature has long discussed the importance of sentiment in vaccination rollouts, this study directly assessed the relationship between sentiment and vaccine uptake. Our findings indicate a weak and somewhat contradictory relationship between vaccine sentiment and vaccination uptake. Although previous research found that positive sentiment (ie, confidence in the vaccine) and uptake increased significantly over time [36], our findings differ in



direction and strength; we found a weak inverse relationship between sentiment and vaccine uptake. Although people in Australia expressed negative sentiments in tweets, many still received vaccines. In contrast, higher rates of positive sentiment were associated with lower observed uptake rates. These associations might point to the role of accessibility and availability. Public sentiment was positive when vaccines were readily accessible; however, a lack of supply, or availability, led to an increase in negative sentiment. This may reflect times when members of the Australian public were unable to be vaccinated, despite their best efforts. Periods of neutral sentiment did not affect vaccine uptake rates. Hence, involving the public and helping them understand the impact of their decisions by emphasizing the importance of vaccination are crucial to ensure higher vaccine uptake rates.

To explain our findings, we must highlight the nature of the expressed sentiments about the vaccine rollout in the collected data; it is important to distinguish between negative sentiment toward the vaccine and negative sentiment toward government actions supporting or inhibiting the vaccination rollout. This is especially important in the Australian context, as the rollout faced many delays and obstacles.

### ***Negative Sentiment Toward the Vaccine Rollout***

This study demonstrated that vaccine uptake increased, while the general sentiment about the vaccination effort was negative. These findings reflect issues surrounding vaccine availability and accessibility. The topic modeling results highlighted that the most frequently discussed topics in tweets were related to the vaccination rollout, not the vaccine itself. In fact, the specific emotions highlighted within the data showed that many tweets from Australia expressed high levels of trust in the vaccine, indicating that most Australians were confident in the vaccine. Negative emotions, such as anger, observed in this study were largely attributed to the role of the government in the rollout rather than vaccine efficacy or safety concerns. A significant drop in negative sentiment and an increase in positive sentiment after news of newly acquired doses arriving in Australia was publicized is further evidence supporting the role of access and availability in public sentiment toward the vaccination rollout. This is also supported by the fact that negative sentiment only significantly increased during events related to vaccine deaths and side effects and returned to similar pre-event sentiment levels when the events had passed. This finding may be specific to the Australian context, given that high vaccine acceptance rates were recorded during the rollout [39]; however, the issue most commonly experienced by members of the public was the inability to receive the vaccine due to limited availability. These findings support those of previous studies, which showed that vaccine sentiment and acceptability are limited indicators of vaccine uptake but other factors (ie, vaccine availability, the ability to take time off work, and crowded vaccination schedules) more strongly influence an individual's decision to be vaccinated [40].

Looking at the Australian context more closely, major shifts in vaccine acceptance and confidence were seen from 2020 to 2021. A study of COVID-19 vaccine hesitancy in 2020 found that 7% of the Australian population were hesitant about

receiving a COVID-19 vaccine and a further 6% would resist the vaccine uptake [41]. However, the uptake rate during the 2021 rollout indicated lower proportions of vaccine hesitancy, with the majority of the population aged  $\geq 12$  years receiving 2 doses of a COVID-19 vaccine by December 30, 2021. Hence the role of vaccine hesitancy toward COVID-19 vaccines may not be as significant in Australia as in other countries. In fact, the vaccine acceptance rate reached 80% during outbreaks of the Delta strain, indicating that the Australian population views the vaccine as a solution for the COVID-19 pandemic [39]. From the start of the Australian COVID-19 vaccination rollout in February 2021, evidence suggested that the main factors affecting vaccine uptake were the availability and accessibility of the vaccine. However, the media continued highlighting vaccine hesitancy as a factor in vaccine uptake, without evidence of high hesitancy rates [39]. A broader examination of general vaccine acceptance (eg, how parents approve of vaccinations for children) shows that the critical elements for high vaccine uptake include vaccine availability and ease of access to vaccination, rather than underlying hesitancy factors [42]. Barriers to vaccination have long included crowded vaccination schedules, slow dissemination of doses, and limited public access to vaccines [43]. The government's failure to clearly disclose supply and access issues has inflated discussions and public perceptions of vaccine hesitancy. Although there is no doubt that vaccine skeptics who refuse to accept and trust globally endorsed scientific evidence around vaccines, including COVID-19 vaccines, exist within the Australian population [3], many other factors influence vaccine uptake rates [39]. Different barriers may arise with different vaccination rollouts and different time points in each rollout. For example, when a new vaccine is created, logistics and approvals may be key barriers, as identified in this study. Other barriers may include poorly timed public health campaigns that have a negative effect on public sentiment. Clear and transparent reporting of all barriers to vaccine uptake, mapped to different vaccine rollouts and time points, would reduce misperceptions in our community.

### **The Influence of Key Events on Sentiment**

This study is the first to examine the effect of vaccine-related events on vaccine sentiment. Events tested within this study included the launch of 4 public health campaigns, lockdowns, antivaccination rallies, news related to vaccine efficacy, and vaccination mandates. Each event showed varying effects on vaccine sentiment.

### ***Public Health Campaigns***

The campaigns launched by the government to encourage the uptake of COVID-19 vaccines resulted in increased negative sentiment, with some events having a more significant influence on sentiment than others. When COVID-19 vaccine-related tweets were closely analyzed, a general frustration was observed within the period of each campaign launch. For example, Twitter users pointed out that the fear-based campaign, showing a young patient fighting for their life in an intensive care unit, was ill-timed because there were no approved vaccines for the age group represented in the campaign (30-40-year-olds) at that time. Multiple paper also criticized this particular campaign, highlighting that inciting fear is not the solution. In a paper

evaluating the campaign's fear appeal, Speight [44] explains that the limitation of this approach is that it neglects people's "capability and opportunity to make the change." Similarly, campaigns with a more positive emotional approach were also met with negative sentiment because many fully vaccinated Australians were still in lockdown, awaiting higher vaccination rates across their state. Hence, an increase in negative sentiment was found around the time these campaigns were launched. In terms of effectiveness, the campaigns did not lead to positive public sentiment, which was the main evaluation measure for the overall vaccine campaign [25]. Therefore, public health messages may have been more effective if they were released at more appropriate times, suggesting the millions of dollars invested in these campaigns would have been better placed in widening vaccine access and availability [45].

### ***Vaccine Acquisition and Approval Related Events***

All events related to increasing vaccine accessibility and availability resulted in significant decreases in negative sentiment and increases in positive sentiment, except when the Pfizer-BioNTech vaccine was approved for children aged 12-15 years old. Similarly, Tavoschi et al. [8] found that the approval of vaccines for children in Italy negatively impacted public sentiment. Excluding when vaccines were approved for children, an increase in positive sentiment was observed when the government increased access to vaccines. Taken together, this examination of different government actions may indicate that actions (eg, securing more vaccines, extending age groups that could be vaccinated) spoke far more loudly to the Australian population than words (eg, public health communication campaigns).

Vaccine efficacy and safety were questioned when cases of blood clots caused by the Oxford-AstraZeneca vaccine were identified in Australia; however, sentiment analysis showed that negative sentiments only increased significantly during these events. The publicized mixed messages and advice related to the Oxford-AstraZeneca vaccine also contributed to the rise in negative sentiment during this period. Danchin and Buttery [39] discussed the government's mixed messaging when the ATAGI revised the approval for the Oxford-AstraZeneca vaccine for certain age groups. People struggled to understand the related risks and benefits, and even health professionals were confused about the approved administration of the Oxford-AstraZeneca vaccine. Public debates of the risks and relative benefits of the Oxford-AstraZeneca vaccine were prominent in this period as politicians and health care professionals took the debate to mainstream media and social media platforms [46]. This shows that vaccination rollouts should include clear and easy-to-understand messages to the public when changes in approvals and recommendations occur to eliminate any spread of misinformation and rumors [47]. This confirms that public knowledge also influences vaccine uptake [48].

### ***Protests, Lockdowns, and Vaccination Mandates***

Protests against COVID-19 public health measures started in different parts of the world in mid-2020 [49]. In Australia, the majority of such events were seen in 2021, when states went into lockdowns and vaccination mandates were established for

certain sectors. Our findings indicate that antivaccination and antilockdown rallies increased negative public sentiment; however, vaccine uptake rates were still increasing. Similar to previous research findings, our data set shows debates between users on Twitter about the agenda behind these protests, the violence against state officials, and conspiracy theories. Our findings support those of Martin and Vanderslott [49], which describe a conflict between personal values (eg, liberal values) and the collective culture with regard to individual choices (eg, I choose not to get vaccinated). Expression of these values on social media led to an increase in negative sentiment, yet vaccination rates also increased. Lockdowns had different effects based on location. This is relevant to the Australian context, given that certain states endured longer lockdowns (ie, Victoria and NSW) than others (ie, South Australia, WA, and QLD). Lockdowns were associated with high uptake rates of the vaccines. Discussions within the collected tweets showed that tweets highlighted the danger of socializing in large crowds without masks prior to achieving the vaccination of a large proportion of the population. This may explain the high vaccination rates during these periods, where being vaccinated was associated with regaining some freedoms [49,50]. Finally, mandating vaccinations for sectors such as aged care did not significantly influence sentiment or uptake. This is similar to the findings of Lee et al [3], which demonstrated that different motivation measures and legislations failed to change the likelihood of vaccine uptake for certain groups. Mandating vaccination may not influence behavior when strong anti- or provaccination beliefs are present [3]. This finding may be explained by the nature of the aged care sector and vaccination requirements. Aged care workers in Australia must receive a set of vaccinations to work in the sector (eg, hepatitis A and B vaccination). Therefore, mandating COVID-19 vaccination for aged care workers did not significantly influence sentiment, as this population is familiar with such requirements.

### **Limitations and Future Research**

This study aimed to extend the understanding of the effect of public sentiment on vaccine uptake and the role of different government measures in changing public sentiment. Certain limitations apply and must be considered when interpreting these findings. First, the study's scope is specific to the Australian vaccination rollout; therefore, the results may not be generalizable. Future research should investigate global sentiment during vaccination rollouts since 2020. Similarly, the data examined for this research are specific to 2021, the time frame of the collected data. Changes in sentiment and vaccine uptake before and after the collected data may be present, including more recent variant outbreaks, such as the Omicron variant outbreak in early 2022. Second, this study used the daily number of administered doses during the study period as a measure of vaccine uptake. However, it must be noted that other measures, such as clicks and traffic to vaccine-booking portals, views, reach, and engagement, were not analyzed and evaluated in the study, limiting our evaluation of public health campaign effectiveness. This can be overcome by future research acquiring clicks and engagement data from agencies running these campaigns [51,52]. Third, the large sample size used in the study may have reduced the impact of random errors, resulting

in a higher probability of finding statistically significant results [53]. Fourth, a time delay may be present for shifts in sentiment and vaccination, as people may take time before taking action. Future research may test sentiment correlation with vaccinations using a distributed lag model [54,55]. Distributed lag models aim to associate outcome variables with time-dependent predictors (eg, getting vaccinated). Finally, we acknowledge that bots may contaminate data collected from social media platforms. Future research should implement a data-cleaning process to eliminate the effect of bot data and ensure data are representative of public opinion [56]. Future research may also use topic modeling to identify repeated text, which would help eliminate news headlines and bot-generated content.

## Conclusion

This study examined the role of public sentiment on COVID-19 vaccine uptake within Australia. The examination of different government actions indicated that actions such as securing more vaccines had a significant effect on alleviating negative sentiment and encouraging vaccine uptake. Our findings indicate that trust in the vaccine was high, contrary to trust in government rollout efforts. Hence, when negative sentiment was prevalent, vaccine uptake remained high. Future efforts for vaccination rollouts should focus on ensuring vaccines are available and accessible to the public through more coordinated supply chain logistics. Further, it is important to align message delivery with supply chain timings and regularly monitor public sentiment and vaccine uptake through objective data collection methods to eliminate biases.

---

## Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Public health vaccine campaigns in Australia.

[DOCX File, 245 KB - [formative\\_v6i9e37775\\_app1.docx](#)]

---

### Multimedia Appendix 2

Stop words used in collecting Tweets from February to October 2021.

[DOCX File, 16 KB - [formative\\_v6i9e37775\\_app2.docx](#)]

---

### Multimedia Appendix 3

Topics extracted from all tweets within the data set.

[DOCX File, 826 KB - [formative\\_v6i9e37775\\_app3.docx](#)]

---

### Multimedia Appendix 4

Topics extracted from positive-sentiment tweets.

[DOCX File, 826 KB - [formative\\_v6i9e37775\\_app4.docx](#)]

---

### Multimedia Appendix 5

Topics extracted from negative-sentiment tweets.

[DOCX File, 823 KB - [formative\\_v6i9e37775\\_app5.docx](#)]

---

### Multimedia Appendix 6

Topics extracted from neutral-sentiment tweets.

[DOCX File, 822 KB - [formative\\_v6i9e37775\\_app6.docx](#)]

---

## References

1. Toro-Ascuy D, Cifuentes-Muñoz N, Avaria A, Pereira-Montecinos C, Cruzat G, Peralta-Arancibia K, et al. Factors influencing the acceptance of covid-19 vaccines in a country with a high vaccination rate. *Vaccines (Basel)* 2022 Apr 25;10(5):681 [FREE Full text] [doi: [10.3390/vaccines10050681](#)] [Medline: [35632437](#)]
2. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: definition, scope and determinants. *Vaccine* 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: [10.1016/j.vaccine.2015.04.036](#)] [Medline: [25896383](#)]
3. Lee D, Rundle-Thiele S, Li G. Motivating seasonal influenza vaccination and cross-promoting covid-19 vaccination: an audience segmentation study among university students. *Vaccines (Basel)* 2021 Nov 25;9(12):1397 [FREE Full text] [doi: [10.3390/vaccines9121397](#)] [Medline: [34960143](#)]
4. Stahl J, Cohen R, Denis F, Gaudelus J, Martinot A, Lery T, et al. The impact of the web and social networks on vaccination. New challenges and opportunities offered to fight against vaccine hesitancy. *Med Mal Infect* 2016 May;46(3):117-122. [doi: [10.1016/j.medmal.2016.02.002](#)] [Medline: [26987960](#)]

5. d'Onofrio A, Manfredi P, Poletti P. The interplay of public intervention and private choices in determining the outcome of vaccination programmes. *PLoS One* 2012 Oct 1;7(10):e45653 [FREE Full text] [doi: [10.1371/journal.pone.0045653](https://doi.org/10.1371/journal.pone.0045653)] [Medline: [23049682](https://pubmed.ncbi.nlm.nih.gov/23049682/)]
6. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on twitter: topic modeling and sentiment analysis. *J Med Internet Res* 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: [10.2196/24435](https://doi.org/10.2196/24435)] [Medline: [34115608](https://pubmed.ncbi.nlm.nih.gov/34115608/)]
7. Liu B. Sentiment analysis and opinion mining. *Synth Lect Hum Lang Technol* 2012 May 23;5(1):1-167. [doi: [10.2200/s00416ed1v01y201204hlt016](https://doi.org/10.2200/s00416ed1v01y201204hlt016)]
8. Tavoschi L, Quattrone F, D'Andrea E, Ducange P, Vabanesi M, Marcelloni F, et al. Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy. *Hum Vaccin Immunother* 2020 May 03;16(5):1062-1069 [FREE Full text] [doi: [10.1080/21645515.2020.1714311](https://doi.org/10.1080/21645515.2020.1714311)] [Medline: [32118519](https://pubmed.ncbi.nlm.nih.gov/32118519/)]
9. World Health Organization. Coronavirus Disease (COVID-19) Pandemic. 2021. URL: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> [accessed 2021-10-12]
10. Hussain A, Ali S, Ahmed M, Hussain S. The anti-vaccination movement: a regression in modern medicine. *Cureus* 2018 Jul 03;10(7):e2919 [FREE Full text] [doi: [10.7759/cureus.2919](https://doi.org/10.7759/cureus.2919)] [Medline: [30186724](https://pubmed.ncbi.nlm.nih.gov/30186724/)]
11. Dubé E, Vivion M, MacDonald NE. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. *Expert Rev Vaccines* 2015 Jan;14(1):99-117. [doi: [10.1586/14760584.2015.964212](https://doi.org/10.1586/14760584.2015.964212)] [Medline: [25373435](https://pubmed.ncbi.nlm.nih.gov/25373435/)]
12. Smith MJ, Marshall GS. Navigating parental vaccine hesitancy. *Pediatr Ann* 2010 Aug;39(8):476-482. [doi: [10.3928/00904481-20100726-05](https://doi.org/10.3928/00904481-20100726-05)] [Medline: [20704143](https://pubmed.ncbi.nlm.nih.gov/20704143/)]
13. Viswanath K, Bekalu M, Dhawan D, Pinnamaneni R, Lang J, McLoud R. Individual and social determinants of COVID-19 vaccine uptake. *BMC Public Health* 2021 Apr 28;21(1):818 [FREE Full text] [doi: [10.1186/s12889-021-10862-1](https://doi.org/10.1186/s12889-021-10862-1)] [Medline: [33910558](https://pubmed.ncbi.nlm.nih.gov/33910558/)]
14. Attwell K, Lake J, Sneddon J, Gerrans P, Blyth C, Lee J. Converting the maybes: crucial for a successful COVID-19 vaccination strategy. *PLoS One* 2021 Jan 20;16(1):e0245907 [FREE Full text] [doi: [10.1371/journal.pone.0245907](https://doi.org/10.1371/journal.pone.0245907)] [Medline: [33471821](https://pubmed.ncbi.nlm.nih.gov/33471821/)]
15. Donzelli G, Palomba G, Federigi I, Aquino F, Cioni L, Verani M, et al. Misinformation on vaccination: a quantitative analysis of YouTube videos. *Hum Vaccin Immunother* 2018 Jul 03;14(7):1654-1659 [FREE Full text] [doi: [10.1080/21645515.2018.1454572](https://doi.org/10.1080/21645515.2018.1454572)] [Medline: [29553872](https://pubmed.ncbi.nlm.nih.gov/29553872/)]
16. Alberici AI, Milesi P. The influence of the internet on the psychosocial predictors of collective action. *J Community Appl Soc Psychol* 2012 Oct 12;23(5):373-388. [doi: [10.1002/casp.2131](https://doi.org/10.1002/casp.2131)]
17. Kwok SWH, Vadde SK, Wang G. Tweet topics and sentiments relating to covid-19 vaccination among Australian Twitter users: machine learning analysis. *J Med Internet Res* 2021 May 19;23(5):e26953 [FREE Full text] [doi: [10.2196/26953](https://doi.org/10.2196/26953)] [Medline: [33886492](https://pubmed.ncbi.nlm.nih.gov/33886492/)]
18. Ting I, Scott N, Palmer A. Untangling Australia's Vaccine Rollout Timetable. 2021. URL: <https://www.abc.net.au/news/2021-05-28/untangling-australia-s-covid-vaccine-rollout-timetable/100156720> [accessed 2021-10-12]
19. BBC News. Australian PM Is Vaccinated as Rollout Begins. 2021. URL: <https://www.bbc.com/news/world-australia-56143277> [accessed 2021-11-11]
20. Department of Health. ATAGI Statement Regarding COVID-19 Vaccines in the Setting of Transmission of the Delta Variant of Concern. 2021. URL: <https://www.health.gov.au/news/atagi-statement-regarding-covid-19-vaccines-in-the-setting-of-transmission-of-the-delta-variant-of-concern> [accessed 2022-08-29]
21. Department of Health. COVID-19 Vaccines. 2021. URL: <https://www.health.gov.au/initiatives-and-programs/covid-19-vaccines> [accessed 2021-10-12]
22. Nicholas J. Australia Still Lags Many Other Countries on Vaccine Rollout—but It's Catching up Fast. 2021. URL: <https://www.theguardian.com/news/datablog/2021/oct/11/australia-still-lags-many-other-countries-on-vaccine-rollout-but-its-catching-up-fast> [accessed 2022-08-29]
23. Department of Health. COVID-19 Vaccine Rollout Update—2 March 2022. 2022. URL: <https://www.health.gov.au/resources/publications/covid-19-vaccine-rollout-update-2-march-2022> [accessed 2022-03-02]
24. Deng T, Ekachai D, Pokrywczynski J. Global covid-19 advertisements: use of informational, transformational and narrative advertising strategies. *Health Commun* 2022 May 22;37(5):628-636. [doi: [10.1080/10410236.2020.1859725](https://doi.org/10.1080/10410236.2020.1859725)] [Medline: [33349050](https://pubmed.ncbi.nlm.nih.gov/33349050/)]
25. Department of Health. Op COVID SHIELD National COVID Vaccine Campaign Plan. 2021. URL: <https://www.health.gov.au/resources/publications/op-covid-shield-national-covid-vaccine-campaign-plan> [accessed 2022-08-29]
26. Correll D. Social Media Statistics Australia—January 2022. 2022. URL: <https://www.socialmedianews.com.au/social-media-statistics-australia-january-2022/> [accessed 2022-02-01]
27. Mehmet M, Roberts R, Nayeem T. Using digital and social media for health promotion: a social marketing approach for addressing co-morbid physical and mental health. *Aust J Rural Health* 2020 Apr 22;28(2):149-158. [doi: [10.1111/ajr.12589](https://doi.org/10.1111/ajr.12589)] [Medline: [31970832](https://pubmed.ncbi.nlm.nih.gov/31970832/)]
28. Batrinca B, Treleaven PC. Social media analytics: a survey of techniques, tools and platforms. *AI Soc* 2014 Jul 26;30(1):89-116. [doi: [10.1007/s00146-014-0549-4](https://doi.org/10.1007/s00146-014-0549-4)]

29. Kaur A, Gupta V. A survey on sentiment analysis and opinion mining techniques. *J Emerg Technol Web Intell* 2013 Nov 01;5(4):367-371. [doi: [10.4304/jetwi.5.4.367-371](https://doi.org/10.4304/jetwi.5.4.367-371)]
30. Halder S. Python Library for the BytesView API. 2021. URL: <https://github.com/algodommmedia/bytesviewapi-python> [accessed 2022-08-29]
31. Jockers ML. An R Package for the Extraction of Sentiment and Sentiment-Based Plot Arcs from Text. 2015. URL: <https://github.com/mjockers/syuzhet> [accessed 2022-08-29]
32. Mohammad S, Turney P. Crowdsourcing a word-emotion association lexicon. *Comput Intell* 2013;29(3):436-465. [doi: [10.1111/j.1467-8640.2012.00460.x](https://doi.org/10.1111/j.1467-8640.2012.00460.x)]
33. Heng LY, Logeswaran R, Marikannan BP. Performance evaluation of analytics models for trends analysis of news. *J Phys: Conf Ser* 2020 Dec 01;1712(1):012021. [doi: [10.1088/1742-6596/1712/1/012021](https://doi.org/10.1088/1742-6596/1712/1/012021)]
34. Mutanga MB, Abayomi A. Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach. *Afr J Sci Technol Innov Dev* 2020 Oct 08;14(1):163-172. [doi: [10.1080/20421338.2020.1817262](https://doi.org/10.1080/20421338.2020.1817262)]
35. Rahman A, Farhana E, Imtiaz N. Snakes in paradise?: Insecure python-related coding practices in stack overflow. 2019 Presented at: Proceedings of the 16th International Conference on Mining Software Repositories (MSR) IEEE Press; May 26-27, 2019; Montreal, Quebec. [doi: [10.1109/msr.2019.00040](https://doi.org/10.1109/msr.2019.00040)]
36. Halbrook M, Gadoth A, Martin-Blais R, Gray AN, Kashani S, Kazan C, et al. Longitudinal assessment of coronavirus disease 2019 vaccine acceptance and uptake among frontline medical workers in Los Angeles, California. *Clin Infect Dis* 2022 Apr 09;74(7):1166-1173. [doi: [10.1093/cid/ciab614](https://doi.org/10.1093/cid/ciab614)] [Medline: [34292319](https://pubmed.ncbi.nlm.nih.gov/34292319/)]
37. Prentice C, Chen J, Stantic B. Timed intervention in COVID-19 and panic buying. *J Retail Consum Serv* 2020 Nov;57:102203. [doi: [10.1016/j.jretconser.2020.102203](https://doi.org/10.1016/j.jretconser.2020.102203)]
38. Andrade C. The limitations of online surveys. *Indian J Psychol Med* 2020 Nov 13;42(6):575-576 [FREE Full text] [doi: [10.1177/0253717620957496](https://doi.org/10.1177/0253717620957496)] [Medline: [33354086](https://pubmed.ncbi.nlm.nih.gov/33354086/)]
39. Danchin M, Buttery J. COVID-19 vaccine hesitancy: a unique set of challenges. *Intern Med J* 2021 Dec;51(12):1987-1989 [FREE Full text] [doi: [10.1111/imj.15599](https://doi.org/10.1111/imj.15599)] [Medline: [34713544](https://pubmed.ncbi.nlm.nih.gov/34713544/)]
40. Rozbroj T, McCaffery K. The importance of addressing social inequalities and targeting the undecided to promote vaccination against COVID-19. *Lancet Reg Health West Pac* 2021 Sep;14:100250 [FREE Full text] [doi: [10.1016/j.lanwpc.2021.100250](https://doi.org/10.1016/j.lanwpc.2021.100250)] [Medline: [34423325](https://pubmed.ncbi.nlm.nih.gov/34423325/)]
41. Edwards B, Biddle N, Gray M, Sollis K. COVID-19 vaccine hesitancy and resistance: correlates in a nationally representative longitudinal survey of the Australian population. *PLoS One* 2021;16(3):e0248892 [FREE Full text] [doi: [10.1371/journal.pone.0248892](https://doi.org/10.1371/journal.pone.0248892)] [Medline: [33760836](https://pubmed.ncbi.nlm.nih.gov/33760836/)]
42. Lawrence GL, Hull BP, MacIntyre CR, McIntyre PB. Reasons for incomplete immunisation among Australian children. A national survey of parents. *Aust Fam Physician* 2004 Jul;33(7):568-571. [Medline: [15301182](https://pubmed.ncbi.nlm.nih.gov/15301182/)]
43. Leask J, Willaby HW, Kaufman J. The big picture in addressing vaccine hesitancy. *Hum Vaccin Immunother* 2014 Nov 19;10(9):2600-2602 [FREE Full text] [doi: [10.4161/hv.29725](https://doi.org/10.4161/hv.29725)] [Medline: [25483479](https://pubmed.ncbi.nlm.nih.gov/25483479/)]
44. Speight J. Why the Federal Government's COVID-19 Fear Appeal to Sydney Residents Won't Work. 2021. URL: <https://theconversation.com/why-the-federal-governments-covid-19-fear-appeal-to-sydney-residents-wont-work-164317> [accessed 2022-08-29]
45. Australian Government. Supporting Australians through COVID-19. 2021. URL: [https://archive.budget.gov.au/2021-22/download/glossy\\_covid19.pdf](https://archive.budget.gov.au/2021-22/download/glossy_covid19.pdf) [accessed 2022-08-29]
46. Isaacs D. The precautionary principle, the AstraZeneca covid-19 vaccine and mixed messaging. *J Paediatr Child Health* 2021 Apr;57(4):472-473 [FREE Full text] [doi: [10.1111/jpc.15468](https://doi.org/10.1111/jpc.15468)] [Medline: [33817870](https://pubmed.ncbi.nlm.nih.gov/33817870/)]
47. Islam MS, Kamal AM, Kabir A, Southern DL, Khan SH, Hasan SMM, et al. COVID-19 vaccine rumors and conspiracy theories: the need for cognitive inoculation against misinformation to improve vaccine adherence. *PLoS One* 2021;16(5):e0251605 [FREE Full text] [doi: [10.1371/journal.pone.0251605](https://doi.org/10.1371/journal.pone.0251605)] [Medline: [33979412](https://pubmed.ncbi.nlm.nih.gov/33979412/)]
48. Zhang J, While AE, Norman IJ. Nurses' knowledge and risk perception towards seasonal influenza and vaccination and their vaccination behaviours: a cross-sectional survey. *Int J Nurs Stud* 2011 Oct;48(10):1281-1289. [doi: [10.1016/j.ijnurstu.2011.03.002](https://doi.org/10.1016/j.ijnurstu.2011.03.002)] [Medline: [21474136](https://pubmed.ncbi.nlm.nih.gov/21474136/)]
49. Martin S, Vanderslott S. "Any idea how fast 'It's just a mask!' can turn into 'It's just a vaccine!'": from mask mandates to vaccine mandates during the COVID-19 pandemic. *Vaccine* 2021 Oct 28;S0264-410X(21)01351-7 [FREE Full text] [doi: [10.1016/j.vaccine.2021.10.031](https://doi.org/10.1016/j.vaccine.2021.10.031)] [Medline: [34823912](https://pubmed.ncbi.nlm.nih.gov/34823912/)]
50. Carlitz RD, Makhura MN. Life under lockdown: illustrating tradeoffs in South Africa's response to COVID-19. *World Dev* 2021 Jan;137:105168 [FREE Full text] [doi: [10.1016/j.worlddev.2020.105168](https://doi.org/10.1016/j.worlddev.2020.105168)] [Medline: [32895596](https://pubmed.ncbi.nlm.nih.gov/32895596/)]
51. These MS, Ronna B, Ott U. P value interpretations and considerations. *J Thorac Dis* 2016 Sep;8(9):E928-E931 [FREE Full text] [doi: [10.21037/jtd.2016.08.16](https://doi.org/10.21037/jtd.2016.08.16)] [Medline: [27747028](https://pubmed.ncbi.nlm.nih.gov/27747028/)]
52. Yousef M, Dietrich T, Rundle-Thiele S. Social advertising effectiveness in driving action: a study of positive, negative and coactive appeals on social media. *Int J Environ Res Public Health* 2021 Jun 01;18(11):5954 [FREE Full text] [doi: [10.3390/ijerph18115954](https://doi.org/10.3390/ijerph18115954)] [Medline: [34206119](https://pubmed.ncbi.nlm.nih.gov/34206119/)]
53. Yousef M, Dietrich T, Rundle - Thiele S, Alhabash S. Emotional appeals effectiveness in enhancing charity digital advertisements. *J Philanthr Mark* 2022 Jul 31:e1763 [FREE Full text] [doi: [10.1002/nvsm.1763](https://doi.org/10.1002/nvsm.1763)]

54. Zanobetti A, Wand MP, Schwartz J, Ryan LM. Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics* 2000 Sep;1(3):279-292. [doi: [10.1093/biostatistics/1.3.279](https://doi.org/10.1093/biostatistics/1.3.279)] [Medline: [12933509](https://pubmed.ncbi.nlm.nih.gov/12933509/)]
55. Bass FM, Clarke DG. Testing distributed lag models of advertising effect. *J Mark Res* 2018 Nov 28;9(3):298-308. [doi: [10.1177/002224377200900309](https://doi.org/10.1177/002224377200900309)]
56. Mehmet M, Callaghan K, Lewis C. Lots of bots or maybe nots: a process for detecting bots in social media research. *Int J Market Res* 2021 Jun 28;63(5):552-559. [doi: [10.1177/14707853211027486](https://doi.org/10.1177/14707853211027486)]

## Abbreviations

**AI:** artificial intelligence  
**API:** application programming interface  
**ATAGI:** Australian Technical Advisory Group on Immunisation  
**LDA:** latent dirichlet allocation  
**NSW:** New South Wales  
**QLD:** Queensland  
**RQ:** research question  
**WA:** Western Australia

*Edited by A Mavragani; submitted 07.03.22; peer-reviewed by S Kwok, M Rajendran; comments to author 18.03.22; revised version received 30.06.22; accepted 23.08.22; published 15.09.22.*

*Please cite as:*

*Yousef M, Dietrich T, Rundle-Thiele S*

*Actions Speak Louder Than Words: Sentiment and Topic Analysis of COVID-19 Vaccination on Twitter and Vaccine Uptake*

*JMIR Form Res* 2022;6(9):e37775

URL: <https://formative.jmir.org/2022/9/e37775>

doi: [10.2196/37775](https://doi.org/10.2196/37775)

PMID: [36007136](https://pubmed.ncbi.nlm.nih.gov/36007136/)

©Murooj Yousef, Timo Dietrich, Sharyn Rundle-Thiele. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 15.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

Original Paper

# Health Information Sourcing and Health Knowledge Quality: Repeated Cross-sectional Survey

Elena Korshakova<sup>1</sup>, MSc; Jesseca K Marsh<sup>2</sup>, PhD; Samantha Kleinberg<sup>1</sup>, PhD

<sup>1</sup>Department of Computer Science, Stevens Institute of Technology, Hoboken, NJ, United States

<sup>2</sup>Department of Psychology, Lehigh University, Bethlehem, PA, United States

**Corresponding Author:**

Samantha Kleinberg, PhD

Department of Computer Science

Stevens Institute of Technology

1 Castle Point on Hudson

Hoboken, NJ, 07030

United States

Phone: 1 201 216 5614

Email: [samantha.kleinberg@stevens.edu](mailto:samantha.kleinberg@stevens.edu)

## Abstract

**Background:** People's health-related knowledge influences health outcomes, as this knowledge may influence whether individuals follow advice from their doctors or public health agencies. Yet, little attention has been paid to where people obtain health information and how these information sources relate to the quality of knowledge.

**Objective:** We aim to discover what information sources people use to learn about health conditions, how these sources relate to the quality of their health knowledge, and how both the number of information sources and health knowledge change over time.

**Methods:** We surveyed 200 different individuals at 12 time points from March through September 2020. At each time point, we elicited participants' knowledge about causes, risk factors, and preventative interventions for 8 viral (Ebola, common cold, COVID-19, Zika) and nonviral (food allergies, amyotrophic lateral sclerosis [ALS], strep throat, stroke) illnesses. Participants were further asked how they learned about each illness and to rate how much they trust various sources of health information.

**Results:** We found that participants used different information sources to obtain health information about common illnesses (food allergies, strep throat, stroke) compared to emerging illnesses (Ebola, common cold, COVID-19, Zika). Participants relied mainly on news media, government agencies, and social media for information about emerging illnesses, while learning about common illnesses from family, friends, and medical professionals. Participants relied on social media for information about COVID-19, with their knowledge accuracy of COVID-19 declining over the course of the pandemic. The number of information sources participants used was positively correlated with health knowledge quality, though there was no relationship with the specific source types consulted.

**Conclusions:** Building on prior work on health information seeking and factors affecting health knowledge, we now find that people systematically consult different types of information sources by illness type and that the number of information sources people use affects the quality of individuals' health knowledge. Interventions to disseminate health information may need to be targeted to where individuals are likely to seek out information, and these information sources differ systematically by illness type.

(*JMIR Form Res* 2022;6(9):e39274) doi:[10.2196/39274](https://doi.org/10.2196/39274)

**KEYWORDS**

health knowledge; health information seeking; information dissemination; COVID-19; online health information; public health; health literacy; social media; information quality; infodemiology

## Introduction

The quality and credibility of information sources people use to find health-related information play a major role in their health outcomes because this information contributes to people's health knowledge. Research has shown that false information can lead to poor treatment choices in many disorders [1], increased worry [2], and, more recently, reduced intentions to receive COVID-19 vaccinations [3]. Although extremely abundant, the content of online health information sources remains mostly unregulated [4], and people often use their own judgment to evaluate these information sources [5,6]. Internet sources are frequently used, though patients also rely on health care professionals and use other sources, such as family, friends, and newspapers, to supplement their information [7]. Further, as awareness of a novel illness grows, people's knowledge and information sources may also change. During the early stages of the COVID-19 pandemic, there was an increase in online information seeking about COVID-19 [8], with a decrease in searches about other health conditions [9]. People searched for information about COVID-19 and its symptoms and visited social media in response to news stories, but these effects were short lived [10,11]. Given the variety in the type and quality of information available, it is crucial to understand the relationship between where people gather their health information and the quality of their health knowledge, and how this may change over time.

We aim to determine (1) what information sources people trust and use to gather health information, (2) the relationship between information sources used and health knowledge quality, and (3) how people's sources of health information change during an evolving situation, the COVID-19 pandemic, and how this relates to the quality of their health knowledge. A better understanding of the link between health information sources and health knowledge quality could allow health care professionals to redirect patients to reliable information sources, and facilitate shared decision-making by providing insight into patients' current understanding. Exploring all of this during a pandemic allows us to understand how the use of information sources may change as knowledge of a disease accrues.

## Background

We first review research on where people obtain health information, before examining people's health knowledge and finally discuss the link between information sources and knowledge.

## Health Information Sourcing

As individuals have access to increasing amounts of health information outside of direct interaction with their doctors, understanding where people obtain this information has become critically important. Models of health information-seeking behavior (HISB) aim to capture the process by which individuals engage in information seeking and how this relates to their health outcomes. Johnson's [12,13] comprehensive model of information seeking (CMIS) has 3 components: antecedents, meaning factors (eg, beliefs) that influence information seeking; information carrier characteristics (eg, content, sources); and the resulting actions. Although this model has been applied

widely, it was originally developed based on information seeking around breast cancer. However, other work has found that people may avoid information about cancer and genetic screening for it [14], especially when they believe themselves at risk [15], suggesting that information seeking related to cancer may differ from other conditions. Another limitation is not accounting for individuals seeking out health information for other people in their lives [16].

Much work around HISB has focused on the information sources individuals use [17] and prefer [18], with particular emphasis on online sources [19] due to their prevalence [20]. Search engines and social media are the primary sources of health information for the US population [4]. Many adults consult the internet before consulting their doctor [21], particularly when they are dissatisfied with their medical care [22]. The most common health search queries fall into 4 main groups: (1) general understanding of a health condition or diagnosis; (2) treatment options, including procedures or medications; (3) information about health professionals and institutions, such as hospitals and pharmacies; and (4) diet and lifestyle information for chronic concerns [23,24]. Among the major search engines, Google was found to provide the most useful and relevant information for health-related searches [25], while Twitter and YouTube were the most popular social media platforms for health information searches [1,4].

One advantage over static information sources (eg, academic and government websites) is that social media platforms allow people to interact and share information with others. However, as experts do not curate these information sources, their quality is unknown, and their use could negatively affect health decisions and behavior [1]. Wikipedia, a free, open source online encyclopedia, is a middle ground, as the information is not vetted by experts but the community attempts to maintain quality and accuracy. Wikipedia is viewed as a reliable source of information by 64% of Americans regardless of the topic choice [26]. Yet, among Wikipedia's more than 14,000 entries on health topics, only 4% were confirmed as high quality [27]. As people rarely check the quality of online health information, they may be unaware of being exposed to low-quality or false information.

Most prior studies on health information seeking focused on online sources, so less is known about how and when people seek out and use offline information sources, such as medical professionals, family members, or news media [7]. Jacobs et al [28] found that most people go to the internet first, with few people seeking information from family, friends, and coworkers. However, that study looked at health information seeking in general, rather than for specific conditions, so it is still an open question whether people's behavior may differ for commonly experienced conditions, such as a cold, compared to emerging diseases, such as Zika virus. Zhang [29] interviewed individuals about specific health incidents, finding factors influencing how people select information sources, but again did not examine differences across conditions or how sources relate to knowledge.

Beyond a preference to look online, where people choose to obtain health information depends on the health topic, information availability, privacy concerns, quality of the



information people expect to find, and health literacy [4,30]. When concerned about privacy and searching for information about conditions where there may be social stigma, people avoid social media and instead use search engines, journals, and books to learn about their health conditions [4]. Some studies have found a correlation between sociodemographic characteristics and patterns of online health information seeking [31,32]. For example, LaValley et al [33] found that younger people tend to use commercial websites (eg, vitamin suppliers who also provide health information to engage people in e-commerce). In contrast, older people tend to use academic websites. Other work has found that overall older adults are less likely to trust online sources [34] and are less successful at using them [35]. Additionally, men use fewer information sources than women and are more concerned about accuracy, comprehensiveness, and ease of access, while women pay more attention to interpretability and ease of understanding [25]. Religious and charitable organizations, which provide access to free or low-cost health care, are common health information sources for vulnerable populations, including ethnic minorities, people with limited English proficiency, rural residents, and immigrants [36]. Overall, prior work has explored the sources used for health information seeking either in specific demographic groups for general information search or for one specific disorder at a time. However, there has been no research that explores within 1 study how information seeking differs across disease types, thus allowing for extrapolating across types of disease. We instead aim to understand whether there are systematic disease-based differences in where people go for health information.

### **Health Knowledge**

The information people gather about health and illness adds to their health knowledge, meaning their beliefs about factors influencing health, causes of disease, and ways to treat and prevent illness [37]. Health knowledge influences how and whether people follow health guidelines [38-40] and when inaccurate can lead to a wide range of adverse health outcomes and delayed medical care [41,42]. Incorrect health knowledge can contribute to choosing a suboptimal treatment and consequently worsen people's conditions, which is particularly risky for acute diseases [43,44].

Models of health information seeking, such as the CMIS, posit that beliefs influence how people seek out health information, but do not examine the process of belief formation. Many of these models are influenced by and incorporate aspects of the Health Belief Model (HBM), which represents the main factors affecting people's health-related behaviors [45]. An alternative model that is more relevant to our study is the Common-Sense Model of Self-Regulation (CSM) [46,47]. In this model, the individual is thought of as (1) actively solving a problem by seeking information, testing hypotheses about their health, and relating their experiences to information they receive; (2) having an illness representation that guides their behaviors; and (3) having their own beliefs that may differ from others' and from medical consensus. This model suggests that experiences, such as an individual's own experiences or those of family or friends, change illness representations. Thus, over the course of a pandemic where people initially have no personal experience

and then gain experience as well as exposure to media, the CSM predicts beliefs will change. The beliefs that have been studied most systematically are those surrounding medication, finding that beliefs about the necessity of medication and reduced concerns are associated with adherence to treatment [48]. However, this work has not been linked to individuals' information-seeking behavior, so it is not known whether the quality of health knowledge is related to the information sources people consult. To close this gap, we aim to conduct an exploratory analysis of how people's knowledge of causes, risk factors, and preventative measures for different illnesses relates to the information sources they use to search for health information.

### **How Health Information Sourcing Affects Health Knowledge**

Although the importance of people's health knowledge is well established, relatively less is known about how that knowledge is influenced by the specific information sources people use. Kealey and Berkman [49] found a correlation between the health information sources participants use and their mental models of cancer. Participants who learned about cancer from the local news had greater ambiguity about cancer prevention, while people who searched for cancer-related information on the internet and in newspapers had less ambiguity [49]. Relying on lay information sources of health information, such as friends and family, was associated with having a higher likelihood of incorrect beliefs about skin cancer [50]. According to both the HBM and the CSM, these beliefs influence behavior. For example, for mental health conditions, people most often report getting information from personal experiences with diagnosed individuals, and these experiences change perceptions of a host of factors, including the causes of disease and treatment preferences (for a review, see [51]). However, the relationship is not necessarily only in 1 direction, as behaviors can influence information seeking, such as anxiety moderating the relationship between care use and information seeking [52].

Information sources themselves have also been linked to behavior, with individuals using print media and interpersonal sources as information sources being more likely to engage in health behaviors [53]. However, far fewer works have examined how information can influence beliefs. Smokers who sought out health information were more likely to intend to quit, but researchers did not find beliefs to mediate this relationship [54].

Although prior work suggests that information matters to whether people understand how to prevent disease, research into the relationship between specific information sources and health knowledge has been limited. Further, we must better understand where people obtain information and how it influences their beliefs across different types of conditions rather than for health in general or only specific concerns. A better understanding of this relationship may aid medical professionals and policymakers who deliver health information and inform the general population. Thus, in this study, we aim to provide a deeper understanding of health information sourcing, trust in those information sources, and implications for health knowledge by investigating the relationship between where

people obtain health information and what they know about various illnesses.

## Methods

### Materials

We selected 8 illnesses, including both viral (Ebola, common cold, COVID-19, Zika) and nonviral (food allergies, amyotrophic lateral sclerosis [ALS], strep throat, stroke) illnesses. These were chosen to include a mix of severities across viral and nonviral illnesses and within nonviral illnesses to have an even split of chronic (food allergies, ALS) and acute (strep throat, stroke) conditions. Given the time frame during which the study was run, we expected COVID-19 to be on many participants' minds. Including other viral illnesses allowed for comparison to other similarly caused health conditions as a baseline against which COVID-19-related responses could be compared.

For each illness, participants first saw 1 screen with the following 3 questions with [ITEM] filled in with the name of the corresponding illness:

- What do you think leads to [ITEM]? (cause question)
- What makes people more or less likely to develop [ITEM]? (risk factor question)
- How can [ITEM] be prevented? (prevention question)

On the next screen, participants were asked to reflect on where they got this information, with the following prompt (information source question):

*Now we'd like you to list all the places where you think you learned any or all of the information you listed about [ITEM]. If you can't remember where you learned something, describe where you think you would have gone to find this information.*

*Some examples of sources include a particular newspaper or magazine, a specific website, personal experience, a medical professional, or from a family member or friend. This is not exhaustive and there may be other sources as well. Please be as specific as possible.*

All responses were free text. After completing these questions for all 8 illnesses, participants rated the trustworthiness of 16 sources of health information, ranging from 1 (not at all trustworthy) to 7 (extremely trustworthy). Participants were also able to indicate whether they were unable to judge a source's trustworthiness. The 16 unique information sources were chosen from places individuals were a priori thought to use for health information (search engines, doctors, WebMD, government health organizations, Wikipedia, public health campaigns, TV news, news websites or newspapers, family, friends), along with information sources mentioned by participants during pilot testing (social media, health and fitness magazines, YouTube, Reddit, medical journals, personal experience). We ran the survey using Qualtrics.

### Ethical Considerations

The data were collected with Institutional Review Board (IRB) approval from the Stevens Institute of Technology (IRB protocol

#2018-003). Before beginning the study, participants read a consent document informing them about the purpose of the study and noting the voluntary nature of participation. Participants provided consent by clicking a response button.

### Procedure

Participants were recruited over 12 time points using Prolific and were compensated US \$4.50 for participation. We restricted the age range as prior work has found that older adults exhibit significantly different information-seeking behavior online [33,55] and different patterns of trust in such sources [34]. Payment was lower during the first time point (US \$3), but we found the study duration to be longer than expected (mean 29.77 minutes, SD 17.21 across all time points) and increased payment accordingly to maintain the target hourly rate. A total of 2350 (97.92%) of 2400 participants remained in analysis, as 4 (0.17%) did not complete the study and 46 (1.91%) were excluded: 17 (37%) responses were random sentences unrelated to the question, 17 (37%) were copied from the internet or the instructions, 7 (15%) were duplicates, 4 (9%) participants reported being aged <18 years or >65 years, and 1 (2%) response was written in Polish.

After consenting to the study, all participants completed the 4 questions (cause, risk factor, prevention, information source) for each illness, with the order of illnesses randomized, followed by the trust ratings for the various predetermined information sources. Following this, participants completed a demographic questionnaire, which, in addition to age, gender, and education level, asked about their state of residence and COVID-19-related restrictions where the participants lived. The survey was run every other week for 20 weeks starting March 31, 2020 (10 time points), and then again 4 weeks later and 2 weeks after that (September 22, 2020) for a total of 12 time points. This frequency was chosen to be often enough to capture changes during the beginning phase of the pandemic. The 2 final time points were chosen to be 1 week before and 2 weeks after Labor Day. These dates coincided with a time in the United States when people often travel and the start of the school year, both events predicted to lead to a surge in COVID-19 cases. Individuals were only allowed to participate in 1 time point.

### Participants

We recruited 200 participants per time point across the 12 time points (2400 total) using Prolific. All participants were US residents aged 18-64 years. [Multimedia Appendix 1](#) includes detailed demographic information about the 2350 participants in our analysis. Across all 12 time points, our sample included 1081 (46%) women, 1222 (52% men), and 47 (2%) who identified in other ways. Participants were mainly younger adults, with 775 (33%) being 18-24, 869 (37%) being 25-34, 400 (17%) being 35-44, 212 (9%) being 45-54, and 94 (4%) being 55-64 years old.

### Data Analysis

All data on causes, risk factors, and preventive measures were coded independently by 2 coders, with disagreements discussed and resolved either by the coders or by a third party. Information sources were coded at a fine level of granularity (eg, biology class) and then mapped to higher-level categories (eg,

education). Responses in which participants indicated that they did not know (I do not know [IDK]) were excluded from analysis at the item level (eg, response to ALS risk factors excluded for a participant coded as IDK, while ALS causes and prevention remained in analysis). For each condition, we determined whether individual responses were correct using the websites of government agencies responsible for international public health, namely the Centers for Disease Control and Prevention (CDC), the World Health Organization, the American Stroke Association, and the Asthma and Allergy Foundation of America.

To evaluate health knowledge quality about causes, risk factors, and preventive measures, we assigned 0 to incorrect responses and 1 to correct ones. See [Multimedia Appendices 2 and 3](#) for a list of all codes used and what responses were considered correct for each illness, [Multimedia Appendix 4](#) for where the correct answers were drawn from, and [Multimedia Appendix 5](#) for an example of how 1 participant's responses were coded. Responses to the cause and risk factor prompts were similar, with participants often listing the same factors for both prompts, so these were combined for analysis. We calculated knowledge precision and knowledge depth (recall) as follows. Knowledge precision is the number of correct unique responses for a participant divided by the total number of unique responses by that participant. For cause and risk factor categories, as long as the response was accurate for either cause or risk factors, it was considered correct. We then averaged precision across cause/risk factor and prevention to yield 1 precision score for each individual for each illness. We calculated knowledge depth by dividing a participant's number of correct responses for an illness and category by the number of codes considered correct for that illness and category. Thus, if the CDC and other sources list A and B as causes or risk factors of COVID-19 and Amy

lists only B, Amy's precision is 1.0 and recall/knowledge depth is 0.5. Together, these measures indicate what fraction of a participant's responses are correct (knowledge precision) and the extent of their knowledge about a condition (knowledge depth).

As our study was exploratory, we conducted a series of analyses to answer key open questions about health information sourcing. We first examined where people obtain information (using prevalence of listed sources) and how sources used differ across illnesses (using a factor analysis). Our factor analysis was run with maximum likelihood estimation using direct oblimin rotation. We used the Scree test and retained all factors with an eigenvalue of 1 or greater. We then examined how much people trust various information sources in general, and then compared average trust in sources an individual used to sources they did not use, tested with the Kolmogorov-Smirnov test. The third key facet we examined is quality of knowledge. We used the coded responses to compare knowledge quality across illnesses and then examined whether knowledge quality is correlated with the number of sources consulted. Finally, we examined how knowledge quality about COVID-19 and the understanding of local restrictions related to the pandemic changed over time.

## Results

### Where Do People Obtain Their Health Information?

We begin with the analysis of the first time point (March 31, 2020) before examining how results changed over time. To determine where people obtain their health information, we report the percentage of individuals mentioning each high-level information source type (eg, news, rather than CNN) for each illness ([Table 1](#)). See [Multimedia Appendix 6](#) for the full list of high-level information source types.

**Table 1.** Top 10 information sources participants used to obtain health information by illness (N=200).

Information source	Food allergies, n (%)	ALS <sup>a</sup> , n (%)	Common cold, n (%)	COVID-19, n (%)	Ebola, n (%)	Strep throat, n (%)	Stroke, n (%)	Zika, n (%)	Total <sup>b</sup> , n (%)
News	16 (8)	20 (10)	16 (8)	66 (33)	70 (35)	6 (3)	10 (5)	66 (33)	36 (18)
Family	24 (12)	10 (5)	30 (15)	12 (6)	8 (4)	30 (15)	36 (18)	8 (4)	20 (10)
Medical professional	20 (10)	4 (2)	24 (12)	6 (3)	2 (1)	38 (19)	18 (9)	6 (3)	14 (7)
Government agency	2 (1)	2 (1)	2 (1)	28 (14)	12 (6)	2 (1)	2 (1)	12 (6)	12 (6)
Social media	4 (2)	16 (8)	4 (2)	18 (9)	14 (7)	2 (1)	4 (2)	10 (5)	12 (6)
Friend	24 (12)	6 (3)	12 (6)	10 (5)	6 (3)	14 (7)	8 (4)	6 (3)	10 (5)
Education	18 (9)	10 (5)	22 (11)	2 (1)	8 (4)	12 (6)	22 (11)	4 (2)	10 (5)
Internet	12 (6)	10 (5)	10 (5)	6 (3)	10 (5)	10 (5)	6 (3)	8 (4)	8 (4)
Website	8 (4)	2 (1)	6 (3)	4 (2)	6 (3)	14 (7)	8 (4)	6 (3)	6 (3)
Personal experience	8 (4)	0	12 (6)	0	0	22 (11)	2 (1)	0	6 (3)

<sup>a</sup>ALS: amyotrophic lateral sclerosis.

<sup>b</sup>Total represents the percentage of participants mentioning each information source when combining all illnesses for analysis of the first time point (March 31, 2020). This value was used for sorting information sources in the table.

We conducted exploratory factor analysis to determine whether there are subgroups or patterns in the information sources used. For all illnesses (N=8), we input the number of mentions for each high-level information source (N=55), leading to an 8×55 matrix where each cell is a count of mentions for that source type for that illness, across all participants. We found that illnesses clustered into 2 distinct groups (Tables 2 and 3), with 4 illnesses loaded ( $\geq .9$ ) on the first factor and 3 other illnesses loaded ( $> .9$ ) on the second factor. One group, which we refer to as “emerging illnesses,” included newly appearing infectious diseases significantly impacting public health (Zika, COVID-19, and Ebola), while the other, “common illnesses,” included those illnesses that historically have been frequent among the general population (stroke, strep throat, food allergies, and common

cold). The most frequently mentioned sources for emerging illnesses were the news (n=68, 34% mentions), government agencies (n=18, 9%), and social media (n=14, 7%). For common illnesses, the most frequently mentioned sources were family (n=30, 15%), medical professionals (n=26, 13%), and friends (n=14, 7%). ALS differed significantly from these 2 groups, reflecting that people learned about it in idiosyncratic ways. Additionally, many participants reported not knowing anything about ALS. When participants did report an information source, it was often related to social media campaigns (eg, ice bucket challenge) or information about celebrities, so this is less likely to reflect intentional information seeking, and ALS was excluded from further analysis.

**Table 2.** Rotated component matrix.

	Factor 1	Factor 2
Common cold	.952	N/A <sup>a</sup>
Strep throat	.941	N/A
Stroke	.915	N/A
Food allergies	.904	.323
Ebola	N/A	.976
Zika	N/A	.972
COVID-19	N/A	.942
ALS <sup>b</sup>	.348	.680

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

<sup>b</sup>ALS: amyotrophic lateral sclerosis.

**Table 3.** Correlation of information sources used between illnesses.

	ALS <sup>a</sup>	Common cold	COVID-19	Ebola	Strep throat	Stroke	Zika
Food allergies	.21	.9	.22	.24	.76	.76	.23
ALS	N/A <sup>b</sup>	.2	.61	.74	-.15	.18	.69
Common cold	N/A	N/A	.14	.18	.84	.87	.17
COVID-19	N/A	N/A	N/A	.94	-.12	.03	.94
Ebola	N/A	N/A	N/A	N/A	-.11	.03	.99
Strep throat	N/A	N/A	N/A	N/A	N/A	.66	-.11
Stroke	N/A	N/A	N/A	N/A	N/A	N/A	.06

<sup>a</sup>ALS: amyotrophic lateral sclerosis.

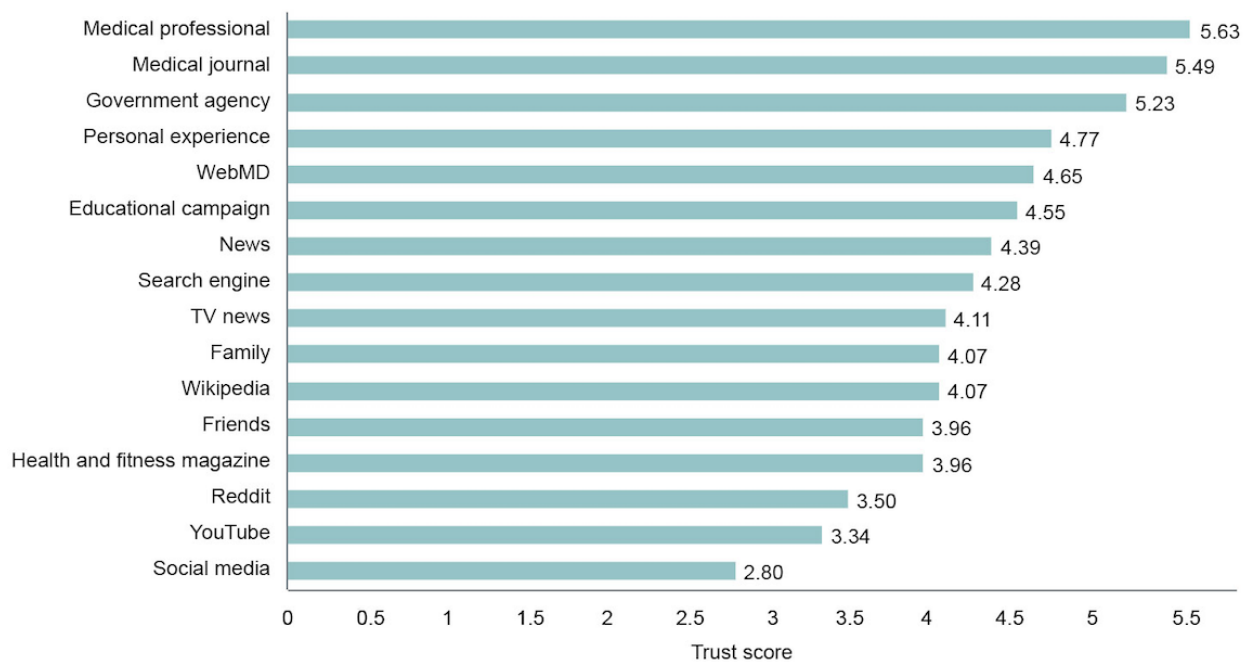
<sup>b</sup>N/A: not applicable.

### Do People Use the Information Sources They Find Trustworthy?

We now examine how participants’ perception of the trustworthiness of information sources relates to those they consult. First, we calculated average trust in the 16 information

sources participants rated at the first time point. Social media received the lowest rating (mean 2.8, SD 1.21), followed by YouTube (mean 3.34, SD 1.23) and Reddit (mean 3.50, SD 1.41), while medical professional (mean 5.63, SD 1.03), medical journal (mean 5.49, SD 1.96), and government agency (mean 5.23, SD 1.41) were rated as the most trustworthy (Figure 1).

**Figure 1.** Average trust score by category for time point 1 (March 31, 2020).



To calculate people’s trust in the information sources they reported using, we mapped the sources each individual used to learn about each illness to the 16 information sources they rated (eg, school nurse was assigned the rating given for medical professional). We then performed the Kolmogorov-Smirnov test to compare the average trust scores for information sources each participant used to the average trust scores for sources they did not use for common and emerging illnesses. The ratings were averaged for unique information sources participants used for each group of illnesses (common and emerging) and separately for sources participants did not use. We found that trust scores were significantly higher for sources participants used to learn about both common illnesses ( $D_{198}=.26, P<.001$ ; used mean 4.32, SD 1.11; not used mean 3.97, SD 1.23) and emerging illnesses ( $D_{198}=.39, P<.001$ ; used mean 4.52, SD 1.09; not used mean 4.03, SD 1.17), meaning people considered the information sources they used more trustworthy than those they did not use. Thus, even if social media is not widely considered a reliable source of health information, compared to individuals who do not use it, individuals who use it believe it is more reliable.

**Are Information Sources Correlated With Knowledge Quality?**

To evaluate participants’ knowledge quality and understand its relationship to information sourcing, we averaged knowledge

precision and knowledge depth separately by participant within each individual illness group (eg, we averaged knowledge precision for COVID-19, Zika, and Ebola to represent emerging illnesses; detailed data by illness are shown in [Multimedia Appendix 7](#)). As the data are not normally distributed (see [Multimedia Appendices 8-11](#)), we performed the Kolmogorov-Smirnov test to compare both knowledge precision and knowledge depth between common and emerging illnesses for statistical significance. We found that knowledge precision for emerging illnesses was significantly higher than for common illnesses ( $D_{198}=.27, P=.003$ ). To examine the extent of participants’ knowledge, we now turn to analysis of knowledge depth ([Table 4](#)). We found again that participants had significantly higher scores for emerging (mean 0.67, SD 0.23) versus common (mean 0.51, SD 0.27) illnesses ( $D_{198}=.19, P<.001$ ). Finally, we performed Spearman rank correlation to test the relationship between the mean number of information sources participants used and the quality of their knowledge. We found that the number of information sources participants reported using was weakly positively correlated with knowledge precision for both common ( $r_{198}=.15, P<.001$ ) and emerging ( $r_{198}=.12, P<.001$ ) illnesses. We observed a stronger correlation between knowledge depth and the number of information sources reported for both common ( $r_{198}=.31, P<.001$ ) and emerging ( $r_{198}=.29, P<.001$ ) illnesses.

**Table 4.** Knowledge precision and knowledge depth for common and emerging illnesses.

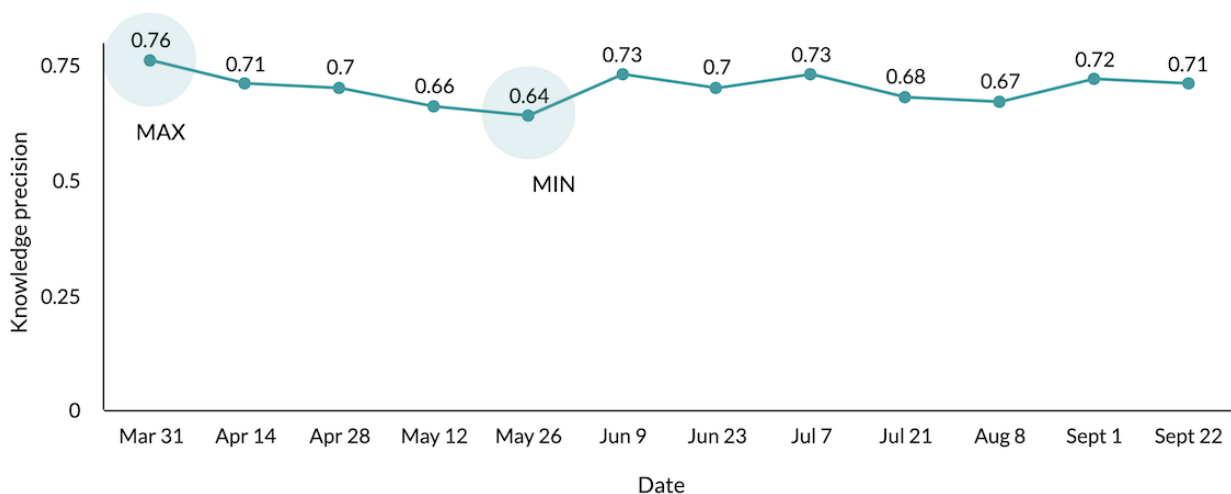
Emerging illnesses		Common illnesses	
Knowledge precision, mean (SD)	Knowledge depth, mean (SD)	Knowledge precision, mean (SD)	Knowledge depth, mean (SD)
0.53 (0.19)	0.67 (0.23)	0.45 (0.18)	0.51 (0.27)

## How Do Knowledge Quality and the Number of Information Sources Change Over Time for COVID-19?

Our second key analysis examined how information sources and trust in sources changed over time from March 31 to September 22, 2020, for COVID-19. We ran a linear trend analysis using the number of information sources participants listed and their trust scores across all 12 time points. We did not find significant changes over time in the number of information sources used ( $F_{1,184}=8.23$ , mean-square error [MSE]=0.16,  $P=.37$ ) or average trust scores across all 16 information sources ( $F_{1,184}=12.76$ , MSE=0.22,  $P=.45$ ), nor did we find a change in trust in any individual source (all  $P>.09$ ). However, we did observe significant changes in knowledge quality using linear trend analysis. The highest knowledge precision score for COVID-19 was found at the start of data collection, on March 31, 2020 (mean 0.76, SD 0.19), when knowledge depth was also high (mean 0.72, SD 0.12). Both knowledge precision and knowledge depth significantly decreased (knowledge precision:  $F_{1,184}=5.31$ , MSE=0.25,  $P=.03$ ;

knowledge depth:  $F_{1,184}=3.02$ , MSE=0.18,  $P<.001$ ) by the last time point, on September 22, 2020 (knowledge precision: mean 0.71, SD 0.22; knowledge depth: mean 0.69, SD 0.18), as shown in Figure 2. Additionally, we measured the effect size, representing the drop in accuracy from one time point to the next using the Cohen  $d$  score. We found that the effect size between time points varied from small to medium ( $d=0.32-0.49$ ). These findings suggest that people were reporting more incorrect information, in addition to correct information, as time went on. Notably, we did not find a significant difference over time for either emerging or common illnesses (excluding COVID-19) for the number of sources used (common:  $F_{1,184}=6.65$ , MSE=0.11,  $P=.19$ ; emerging:  $F_{1,184}=7.89$ , MSE=0.14,  $P=.09$ ), trust in sources (common:  $F_{1,184}=9.86$ , MSE=0.15,  $P=.49$ ; emerging:  $F_{1,184}=14.21$ , MSE=0.09,  $P=.77$ ), knowledge precision (common:  $F_{1,184}=10.31$ , MSE=0.13,  $P=.08$ ; emerging:  $F_{1,184}=13.64$ , MSE=0.11,  $P=.35$ ) or knowledge depth (common:  $F_{1,184}=11.73$ , MSE=0.14,  $P=.21$ ; emerging:  $F_{1,184}=8.67$ , MSE=0.08,  $P=.45$ ).

**Figure 2.** Knowledge precision (combining causes, risk factors, and prevention) about COVID-19 across all 12 time points.



Similarly, we examined how knowledge about COVID-19-related restrictions changed over time to find out how aware participants were of rapidly changing health guidelines. We collapsed all states for analysis (after finding no significant difference in scores between regions of the United States, nor based on state-level vaccination rates, which could indicate differences in COVID-19-related beliefs). Linear trend analysis showed a decline in both knowledge precision ( $F_{1,184}=12.91$ , MSE=0.19,  $P=.003$ ) and knowledge depth ( $F_{1,184}=8.004$ , MSE=0.26,  $P<.001$ ) of COVID-19-related restrictions. Additionally, we examined the types of incorrect responses by comparing how often people thought there were more restrictions versus fewer. To compare these scores, we calculated the number of incorrect responses by participant within these 2 categories and compared them for statistical significance using the Kolmogorov-Smirnov test. Incorrect responses were mainly due to participants being unaware of restrictions in their state and believing there were fewer restrictions than were actually imposed ( $D_{1649}=0.23$ ,  $P<.001$ ).

## Discussion

### Principal Findings

In this study, we aimed to learn what information sources people use to learn about health conditions, how these sources relate to the quality of their health knowledge, and how both information source choices and health knowledge change over time. First, we found that people systematically rely on different types of sources, depending on whether they are obtaining information about common illnesses (food allergies, strep throat, stroke) or emerging illnesses (Ebola, common cold, COVID-19, Zika), with individuals relying on the news, government agencies, and social media for gathering information about emerging illnesses, while seeking information from family, friends, and medical professionals for common illnesses. Second, we found that the number of sources people use is positively correlated with the quality of health knowledge. Interestingly, participants had a better understanding of emerging illnesses than common ones, which may be explained

by the fact that participants were more likely to use lay information sources, such as friends and family, to learn about common illnesses, suggesting those information sources are more likely to be incomplete or incorrect.

Although participants overall did distinguish between the trustworthiness of different information sources, they rated sources they used as more trustworthy than those they did not use. For example, although social media received low trust scores overall, compared to individuals who do not use it as a source of health information, those who do find it more trustworthy. Finally, throughout 6 months of the early stage of the COVID-19 pandemic, we found that knowledge quality overall declined, including knowledge of causes, risk factors, and preventative interventions, along with knowledge of local COVID-19-related restrictions. We did not find any change over time for the other conditions, which suggests that despite our cross-sectional approach, changes in knowledge over time were specific to COVID-19 and not due to population differences. This decrease in knowledge may be in part because over the course of the pandemic, the number of different, nuanced, restrictions grew. As the complexity in restrictions grew, it appears our participants were less able to keep track of what was in place in their state. Future work should explore how the complexity of public health restrictions relates to the knowledge of those restrictions and how these findings can be used to develop guidelines on information dissemination for emerging illnesses.

### Comparison With Prior Work

Our study can be understood in the context of the CSM [47], whereby illness representations, which include knowledge of causes and control (eg, preventive strategies, treatment), guide behavior. Building on this framework, we now find that people's knowledge of both differs systematically between common and emerging illnesses as do their information sources. Although the link between beliefs and specific source choices has not been examined, prior work has framed information search as a hypothesis-testing process, suggesting that beliefs may guide what people search for online [56]. It remains for future work to determine whether information quality (eg, use of lay and interpersonal sources) drives reduced knowledge quality for common illnesses or alternatively whether personal factors independently drive both knowledge and source selection. The CSM also provides a framework for interpreting our findings on COVID-19. In addition to the growing complexity of COVID-19 knowledge and restrictions over time (eg, just stay home vs guidelines on masking, isolation after infection, and ventilation), people may also perceive the health threat differently over time as they gain personal experience or find that it is increasingly well understood. This in turn may influence their information sources or information-seeking behavior. Future work during emerging pandemics is needed to examine the relationship between information sources, beliefs, and perceived risk.

We also built on prior work on health information seeking by identifying systematic differences across illnesses. Previous studies have focused primarily on the use of online health information sources, such as social media and websites

[1,2,4,31], and people's knowledge about specific illnesses [5,21,22] rather than finding patterns in illness groups. Instead of focusing on specific information sources, we used open-ended questions collected across a range of times, enabling us to learn both where people obtain their information and how this may change. Our findings regarding the most common health information sources are aligned with prior work investigating health information-seeking strategies [1,2,4]; however, prior studies did not investigate search patterns by illness group and their correlation with health knowledge. Differences in the sources of information that people use for emerging and common illnesses may be explained by the nature of the illness. Common illnesses, such as food allergies, the common cold, strep throat, and stroke, may be learned about from family members over the course of a lifetime as they affect individuals or those they know. When a new disorder emerges, individuals do not have personal experience and they rely on news media and government agencies, both of which disseminate information about novel illnesses [57]. This difference provides an opportunity to extend our findings to other illnesses that may be categorized as either common or emerging diseases and to develop effective information delivery strategies.

We further found that although overall, people rate medical professionals and government sources of information as more trustworthy than lay sources (eg, social media, friends), compared to individuals who do not use lay information sources for obtaining health information, people who choose to do so rate those sources as more trustworthy. This has 2 key implications. First, there is a need to help individuals access reliable information and to evaluate the quality of information they receive. People may be at risk of making wrong decisions because of misconceptions regarding their health and lack of awareness of the quality of the health information they use. This could contribute to inappropriate health interventions, misinterpretation of new information, and poor health outcomes. Second, knowing where individuals are looking for information during a health emergency, such as a pandemic, versus when they have a common condition means that information can be selectively targeted to where individuals will look for it. For example, Bautista et al [58] proposed having health professionals correct misinformation on social media. Our results provide further support for this. Wider health communication campaigns on social media can focus on using medical professionals to provide reliable information about emerging illnesses, such as the ongoing coronavirus pandemic. Knowing that participants commonly use social media even though they recognize it as less trustworthy can be leveraged by government agencies to more effectively deliver information, such as about vaccination, and potentially improve health knowledge.

### Limitations

Since our data were collected from Prolific and only from US residents, our study may not generalize to other populations. Thus, future research may be needed to investigate the relationship between information sources and health knowledge quality in geographically diverse samples. The nature of the illnesses studied is also such that we cannot learn in real time where people obtained information (eg, when they first heard about strep throat), so there may be bias due to relying on

participants' memory and surveying different participants over time. As we did not ask participants whether they had any of the listed conditions, it is possible that personal experience may be a moderator of information sources or knowledge quality. Using multiple categories of illnesses allows us to see that participants are not simply listing the sources they used for the most recent emerging illness (COVID-19), but we cannot know whether an individual's specific information sources or knowledge quality are changing. Our analyses were mainly exploratory, though conducting many analyses could introduce  $\alpha$  errors. Further, although multiple coders independently coded each free-text response, this qualitative analysis process could potentially introduce bias. Finally, we solicited ratings of trust for 16 categories of information sources, but this did not cover more infrequently mentioned source types and combined categories of sources (eg, family, without distinguishing between people in a caregiving role or by closeness of the relationship).

Other key avenues for future research include examining trust at a more granular level (eg, the specific sources each participant listed), changes in belief and information seeking at the individual level (eg, to discover whether an individual may shift from social media to family for information as an illness becomes endemic in a population), how using an information source for an emerging disease may change information source

use for previously known diseases, and how these findings can be used to deliver information more effectively. Although we know what misconceptions people may have and where they are seeking information, future work is needed to translate this into effective message strategies.

## Conclusion

Our study investigated what information sources people use to learn health information and how these sources are related to individuals' knowledge. The results of our study indicate that health information sources and knowledge quality vary by illness type. We identified 2 groups of illnesses according to information sources people consult to find health information: common illnesses (food allergies, strep throat, stroke) that are mainly learned from family, friends, and medical professionals and emerging illnesses (Ebola, common cold, COVID-19, Zika) that people learn about from news media, government agencies, and social media. We also found that people tend to trust information sources they use even when most people consider them low quality (eg, social media). Our findings suggest a need for targeted interventions via information sources people are likely to use for different illness types. In the future, we aim to extend this study by investigating how incorrect knowledge affects the way people use information and make health-related decisions.

---

## Acknowledgments

We thank Helen Keetley, Hillary Winoker, Raihan Alam, Marissa Alverzo, John Brummer, Amanda Klansky, Taylor Kurz, Haley Miller, Renali Patel, Siddanth Patel, Gabriella Puchall, Sophia Salomoni, Madison Schulte, Boris Tzankov, and Nada Younies for assistance in annotating data. This work was supported in part by the James S. McDonnell Foundation and the National Science Foundation (NSF; award #1915182).

---

## Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Percentage distribution of participants by demographic.

[\[DOCX File, 18 KB - formative\\_v6i9e39274\\_app1.docx\]](#)

---

### Multimedia Appendix 2

List of codes for causes and risk factors and correctness by illness.

[\[DOCX File, 23 KB - formative\\_v6i9e39274\\_app2.docx\]](#)

---

### Multimedia Appendix 3

List of codes for preventative measures and correctness by illness.

[\[DOCX File, 19 KB - formative\\_v6i9e39274\\_app3.docx\]](#)

---

### Multimedia Appendix 4

List of sources for assigning the correctness of participants' responses.

[\[DOCX File, 14 KB - formative\\_v6i9e39274\\_app4.docx\]](#)

---

### Multimedia Appendix 5

Coding protocol sample.

[\[DOCX File, 17 KB - formative\\_v6i9e39274\\_app5.docx\]](#)

---



## Multimedia Appendix 6

List of all high-level information sources used for categorizing participant responses.

[[DOCX File , 13 KB - formative\\_v6i9e39274\\_app6.docx](#) ]

---

## Multimedia Appendix 7

Knowledge precision and depth across the illnesses on March 31 (timepoint 1).

[[DOCX File , 14 KB - formative\\_v6i9e39274\\_app7.docx](#) ]

---

## Multimedia Appendix 8

Knowledge precision distribution for emerging illnesses.

[[PNG File , 62 KB - formative\\_v6i9e39274\\_app8.png](#) ]

---

## Multimedia Appendix 9

Knowledge precision distribution for common illnesses.

[[PNG File , 66 KB - formative\\_v6i9e39274\\_app9.png](#) ]

---

## Multimedia Appendix 10

Knowledge depth distribution for emerging illnesses.

[[PNG File , 64 KB - formative\\_v6i9e39274\\_app10.png](#) ]

---

## Multimedia Appendix 11

Knowledge depth distribution for common illnesses.

[[PNG File , 59 KB - formative\\_v6i9e39274\\_app11.png](#) ]

---

## References

1. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. *Health Inform J* 2015 Sep;21(3):173-194 [[FREE Full text](#)] [doi: [10.1177/1460458213512220](https://doi.org/10.1177/1460458213512220)] [Medline: [24670899](https://pubmed.ncbi.nlm.nih.gov/24670899/)]
2. Liu PL. COVID-19 information seeking on digital media and preventive behaviors: the mediation role of worry. *Cyberpsychol Behav Soc Netw* 2020 Oct;23(10):677-682. [doi: [10.1089/cyber.2020.0250](https://doi.org/10.1089/cyber.2020.0250)] [Medline: [32498549](https://pubmed.ncbi.nlm.nih.gov/32498549/)]
3. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat Hum Behav* 2021 Mar;5(3):337-348. [doi: [10.1038/s41562-021-01056-1](https://doi.org/10.1038/s41562-021-01056-1)] [Medline: [33547453](https://pubmed.ncbi.nlm.nih.gov/33547453/)]
4. De Choudhury M, Morris MR, White RW. Seeking and sharing health information online: comparing search engines and social media. 2014 Presented at: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, ON p. 1365-1376. [doi: [10.1145/2556288.2557214](https://doi.org/10.1145/2556288.2557214)]
5. 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](https://doi.org/10.1016/j.socscimed.2007.01.012)] [Medline: [17328998](https://pubmed.ncbi.nlm.nih.gov/17328998/)]
6. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. *J Med Internet Res* 2019 May 02;21(5):e12522 [[FREE Full text](#)] [doi: [10.2196/12522](https://doi.org/10.2196/12522)] [Medline: [31045507](https://pubmed.ncbi.nlm.nih.gov/31045507/)]
7. Cutilli CC. Seeking health information: what sources do your patients use? *Orthop Nurs* 2010;29(3):214-219. [doi: [10.1097/NOR.0b013e3181db5471](https://doi.org/10.1097/NOR.0b013e3181db5471)] [Medline: [20505493](https://pubmed.ncbi.nlm.nih.gov/20505493/)]
8. Jokic-Begic N, Lauri Korajlija A, Mikac U. Cyberchondria in the age of COVID-19. *PLoS One* 2020;15(12):e0243704 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0243704](https://doi.org/10.1371/journal.pone.0243704)] [Medline: [33332400](https://pubmed.ncbi.nlm.nih.gov/33332400/)]
9. Mangono T, Smittenaar P, Caplan Y, Huang VS, Sutermeister S, Kemp H, et al. Information-seeking patterns during the COVID-19 pandemic across the United States: longitudinal analysis of google trends data. *J Med Internet Res* 2021 May 03;23(5):e22933 [[FREE Full text](#)] [doi: [10.2196/22933](https://doi.org/10.2196/22933)] [Medline: [33878015](https://pubmed.ncbi.nlm.nih.gov/33878015/)]
10. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. *Proc Natl Acad Sci* 2020 May 26;117(21):11220-11222 [[FREE Full text](#)] [doi: [10.1073/pnas.2005335117](https://doi.org/10.1073/pnas.2005335117)] [Medline: [32366658](https://pubmed.ncbi.nlm.nih.gov/32366658/)]
11. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: mixed-methods analysis. *J Med Internet Res* 2020 Oct 12;22(10):e21597 [[FREE Full text](#)] [doi: [10.2196/21597](https://doi.org/10.2196/21597)] [Medline: [32960775](https://pubmed.ncbi.nlm.nih.gov/32960775/)]
12. Johnson J, Meischke H. A comprehensive model of cancer-related information seeking applied to magazines. *Human Commun Res* 1993 Mar;19(3):343-367. [doi: [10.1111/j.1468-2958.1993.tb00305.x](https://doi.org/10.1111/j.1468-2958.1993.tb00305.x)]
13. Johnson JD. *Cancer-Related Information Seeking*. New York, NY: Hampton Press; 1997.

14. Case DO, Andrews JE, Johnson JD, Allard SL. Avoiding versus seeking: the relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. *J Med Libr Assoc* 2005 Jul;93(3):353-362 [[FREE Full text](#)] [Medline: [16059425](#)]
15. Brashers DE, Goldsmith DJ, Hsieh E. Information seeking and avoiding in health contexts. *Human Commun Res* 2002 Apr;28(2):258-271. [doi: [10.1111/j.1468-2958.2002.tb00807.x](#)]
16. Reifegerste D, Blech S, Dechant P. Understanding information seeking about the health of others: applying the comprehensive model of information seeking to proxy online health information seeking. *J Health Commun* 2020 Feb 01;25(2):126-135. [doi: [10.1080/10810730.2020.1716280](#)] [Medline: [32009552](#)]
17. Anker AE, Reinhart AM, Feeley TH. Health information seeking: a review of measures and methods. *Patient Educ Couns* 2011 Mar;82(3):346-354. [doi: [10.1016/j.pec.2010.12.008](#)] [Medline: [21239134](#)]
18. Zimmerman MS, Shaw G. Health information seeking behaviour: a concept analysis. *Health Info Libr J* 2020 Sep;37(3):173-191. [doi: [10.1111/hir.12287](#)] [Medline: [32052549](#)]
19. Jia X, Pang Y, Liu LS. Online Health Information Seeking Behavior: A Systematic Review. *Healthcare* 2021 Dec 16;9(12):1740 [[FREE Full text](#)] [doi: [10.3390/healthcare9121740](#)] [Medline: [34946466](#)]
20. Cline RJ, Haynes KM. Consumer health information seeking on the internet: the state of the art. *Health Educ Res* 2001 Dec;16(6):671-692. [doi: [10.1093/her/16.6.671](#)] [Medline: [11780707](#)]
21. Swoboda CM, Van Hulle JM, McAlearney AS, Huerta TR. Odds of talking to healthcare providers as the initial source of healthcare information: updated cross-sectional results from the Health Information National Trends Survey (HINTS). *BMC Fam Pract* 2018 Aug 29;19(1):146 [[FREE Full text](#)] [doi: [10.1186/s12875-018-0805-7](#)] [Medline: [30157770](#)]
22. Tustin N. The role of patient satisfaction in online health information seeking. *J Health Commun* 2010 Jan;15(1):3-17. [doi: [10.1080/10810730903465491](#)] [Medline: [20390974](#)]
23. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using Internet search engines to obtain medical information: a comparative study. *J Med Internet Res* 2012 May 16;14(3):e74 [[FREE Full text](#)] [doi: [10.2196/jmir.1943](#)] [Medline: [22672889](#)]
24. 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](#)]
25. Rowley J, Johnson F, Sbaffi L. Gender as an influencer of online health information-seeking and evaluation behavior. *J Assoc Inf Sci Technol* 2015 Aug 25;68(1):36-47. [doi: [10.1002/asi.23597](#)]
26. Javanmardi S, Ganjisaffar Y, Lopes C, Baldi P. User contribution and trust in Wikipedia. 2009 Presented at: 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing; November 11-14, 2009; Crystal City, WA. [doi: [10.4108/icst.collaboratecom2009.8376](#)]
27. Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? *J Am Med Inform Assoc* 2009;16(4):471-479 [[FREE Full text](#)] [doi: [10.1197/jamia.M3059](#)] [Medline: [19390105](#)]
28. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. *Cogent Soc Sci* 2017 Mar 13;3(1):1302785. [doi: [10.1080/23311886.2017.1302785](#)]
29. Zhang Y. Beyond quality and accessibility: source selection in consumer health information searching. *J Assoc Inf Sci Technol* 2014 Jan 07;65(5):911-927. [doi: [10.1002/asi.23023](#)]
30. Diviani N, Zanini C, Jaks R, Brach M, Gemperli A, Rubinelli S. Information seeking behavior and perceived health literacy of family caregivers of persons living with a chronic condition. The case of spinal cord injury in Switzerland. *Patient Educ Couns* 2020 Aug;103(8):1531-1537. [doi: [10.1016/j.pec.2020.02.024](#)] [Medline: [32098740](#)]
31. Ali SH, Foreman J, Tozan Y, Capasso A, Jones AM, DiClemente RJ. Trends and predictors of COVID-19 information sources and their relationship with knowledge and beliefs related to the pandemic: nationwide cross-sectional study. *JMIR Public Health Surveill* 2020 Oct 08;6(4):e21071 [[FREE Full text](#)] [doi: [10.2196/21071](#)] [Medline: [32936775](#)]
32. Perez SL, Kravitz RL, Bell RA, Chan MS, Paterniti DA. Characterizing internet health information seeking strategies by socioeconomic status: a mixed methods approach. *BMC Med Inform Decis Mak* 2016 Aug 09;16:107 [[FREE Full text](#)] [doi: [10.1186/s12911-016-0344-x](#)] [Medline: [27506607](#)]
33. LaValley SA, Kiviniemi MT, Gage-Bouchard EA. Where people look for online health information. *Health Info Libr J* 2017 Jun 21;34(2):146-155 [[FREE Full text](#)] [doi: [10.1111/hir.12143](#)] [Medline: [27207817](#)]
34. Miller LMS, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. *J Aging Health* 2012 Apr;24(3):525-541. [doi: [10.1177/0898264311428167](#)] [Medline: [22187092](#)]
35. Agree EM, King AC, Castro CM, Wiley A, Borzekowski DLG. "It's got to be on this page": age and cognitive style in a study of online health information seeking. *J Med Internet Res* 2015 Mar 24;17(3):e79 [[FREE Full text](#)] [doi: [10.2196/jmir.3352](#)] [Medline: [25831483](#)]
36. Wheldon CW, Carroll KT, Moser RP. Trust in health information sources among underserved and vulnerable populations in the U.S. *J Health Care Poor Underserved* 2020;31(3):1471-1487. [doi: [10.1353/hpu.2020.0106](#)] [Medline: [33416706](#)]
37. Reynolds R, Dennis S, Hasan I, Slewa J, Chen W, Tian D, et al. A systematic review of chronic disease management interventions in primary care. *BMC Fam Pract* 2018 Jan 09;19(1):11 [[FREE Full text](#)] [doi: [10.1186/s12875-017-0692-3](#)] [Medline: [29316889](#)]
38. Marsh JK, Ungson ND, Packer DJ. Bring out your experts: the relationship between perceived expert causal understanding and pandemic behaviors. *J Exp Psychol Appl* 2021 Dec;27(4):785-802. [doi: [10.1037/xap0000402](#)] [Medline: [35073134](#)]

39. Brewer NT, Chapman GB, Gibbons FX, Gerrard M, McCaul KD, Weinstein ND. Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. *Health Psychol* 2007 Mar;26(2):136-145. [doi: [10.1037/0278-6133.26.2.136](https://doi.org/10.1037/0278-6133.26.2.136)] [Medline: [17385964](https://pubmed.ncbi.nlm.nih.gov/17385964/)]
40. Bruine de Bruin W, Bennett D. Relationships between initial COVID-19 risk perceptions and protective health behaviors: a national survey. *Am J Prev Med* 2020 Aug;59(2):157-167 [FREE Full text] [doi: [10.1016/j.amepre.2020.05.001](https://doi.org/10.1016/j.amepre.2020.05.001)] [Medline: [32576418](https://pubmed.ncbi.nlm.nih.gov/32576418/)]
41. Livi S, Zeri F, Baroni R. Health beliefs affect the correct replacement of daily disposable contact lenses: predicting compliance with the Health Belief Model and the Theory of Planned Behaviour. *Cont Lens Anterior Eye* 2017 Feb;40(1):25-32. [doi: [10.1016/j.clae.2016.09.003](https://doi.org/10.1016/j.clae.2016.09.003)] [Medline: [27671981](https://pubmed.ncbi.nlm.nih.gov/27671981/)]
42. Vitriol JA, Marsh JK. A pandemic of misbelief: how beliefs promote or undermine COVID-19 mitigation. *Front Polit Sci* 2021 Jun 17;3:54. [doi: [10.3389/fpos.2021.648082](https://doi.org/10.3389/fpos.2021.648082)]
43. Prasanna SS, Korner-Bitensky N, Ahmed S. Why do people delay accessing health care for knee osteoarthritis? Exploring beliefs of health professionals and lay people. *Physiother Can* 2013;65(1):56-63 [FREE Full text] [doi: [10.3138/ptc.2011-50](https://doi.org/10.3138/ptc.2011-50)] [Medline: [24381383](https://pubmed.ncbi.nlm.nih.gov/24381383/)]
44. Jurkowski JM, Maniccia DM, Spicer DA, Dennison BA. Impact of a multimedia campaign to increase intention to call 9-1-1 for stroke symptoms, upstate New York, 2006-2007. *Prev Chronic Dis* 2010 Mar;7(2):A35 [FREE Full text] [Medline: [20158963](https://pubmed.ncbi.nlm.nih.gov/20158963/)]
45. Jones CL, Jensen JD, Scherr CL, Brown NR, Christy K, Weaver J. The Health Belief Model as an explanatory framework in communication research: exploring parallel, serial, and moderated mediation. *Health Commun* 2015;30(6):566-576 [FREE Full text] [doi: [10.1080/10410236.2013.873363](https://doi.org/10.1080/10410236.2013.873363)] [Medline: [25010519](https://pubmed.ncbi.nlm.nih.gov/25010519/)]
46. Leventhal H, Phillips LA, Burns E. The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management. *J Behav Med* 2016 Dec;39(6):935-946. [doi: [10.1007/s10865-016-9782-2](https://doi.org/10.1007/s10865-016-9782-2)] [Medline: [27515801](https://pubmed.ncbi.nlm.nih.gov/27515801/)]
47. Leventhal H, Brissette I, Leventhal EA. The common-sense model of self-regulation of health and illness. In: Cameron LD, Leventhal H, editors. *The Self-Regulation of Health and Illness Behaviour*. Oxfordshire, UK: Routledge; 2003:42-65.
48. Horne R, Chapman SCE, Parham R, Freemantle N, Forbes A, Cooper V. Understanding patients' adherence-related beliefs about medicines prescribed for long-term conditions: a meta-analytic review of the Necessity-Concerns Framework. *PLoS One* 2013;8(12):e80633 [FREE Full text] [doi: [10.1371/journal.pone.0080633](https://doi.org/10.1371/journal.pone.0080633)] [Medline: [24312488](https://pubmed.ncbi.nlm.nih.gov/24312488/)]
49. Kealey E, Berkman CS. The relationship between health information sources and mental models of cancer: findings from the 2005 Health Information National Trends Survey. *J Health Commun* 2010 Dec;15 Suppl 3:236-251. [doi: [10.1080/10810730.2010.522693](https://doi.org/10.1080/10810730.2010.522693)] [Medline: [21154096](https://pubmed.ncbi.nlm.nih.gov/21154096/)]
50. Ford BM, Kaphingst KA. Lay interpersonal sources for health information related to beliefs about the modifiability of cancer risk. *Cancer Causes Control* 2009 Dec;20(10):1975-1983 [FREE Full text] [doi: [10.1007/s10552-009-9392-1](https://doi.org/10.1007/s10552-009-9392-1)] [Medline: [19578935](https://pubmed.ncbi.nlm.nih.gov/19578935/)]
51. Jorm AF. Mental health literacy. Public knowledge and beliefs about mental disorders. *Br J Psychiatry* 2000 Nov;177:396-401. [doi: [10.1192/bjp.177.5.396](https://doi.org/10.1192/bjp.177.5.396)] [Medline: [11059991](https://pubmed.ncbi.nlm.nih.gov/11059991/)]
52. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. *Cyberpsychol Behav* 2006 Aug;9(4):494-498. [doi: [10.1089/cpb.2006.9.494](https://doi.org/10.1089/cpb.2006.9.494)] [Medline: [16901253](https://pubmed.ncbi.nlm.nih.gov/16901253/)]
53. Redmond N, Baer HJ, Clark CR, Lipsitz S, Hicks LS. Sources of health information related to preventive health behaviors in a national study. *Am J Prev Med* 2010 Jun;38(6):620-627.e2 [FREE Full text] [doi: [10.1016/j.amepre.2010.03.001](https://doi.org/10.1016/j.amepre.2010.03.001)] [Medline: [20494238](https://pubmed.ncbi.nlm.nih.gov/20494238/)]
54. Upadhyay S, Lord J, Gakh M. Health-information seeking and intention to quit smoking: do health beliefs have a mediating role? *Tob Use Insights* 2019;12:1179173X19871310 [FREE Full text] [doi: [10.1177/1179173X19871310](https://doi.org/10.1177/1179173X19871310)] [Medline: [31488954](https://pubmed.ncbi.nlm.nih.gov/31488954/)]
55. Sharit J, Hernández MA, Czaja SJ, Pirolli P. Investigating the roles of knowledge and cognitive abilities in older adult information seeking on the web. *ACM Trans Comput Hum Interact* 2008;15(1):3 [FREE Full text] [doi: [10.1145/1352782.1352785](https://doi.org/10.1145/1352782.1352785)] [Medline: [20011130](https://pubmed.ncbi.nlm.nih.gov/20011130/)]
56. Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. *J Am Med Inform Assoc* 2008 Jul;15(4):484-495 [FREE Full text] [doi: [10.1197/jamia.M2449](https://doi.org/10.1197/jamia.M2449)] [Medline: [18436912](https://pubmed.ncbi.nlm.nih.gov/18436912/)]
57. Green MH. Emerging diseases, re-emerging histories. *Centaurus* 2020 Jul 27;62(2):234-247. [doi: [10.1111/1600-0498.12306](https://doi.org/10.1111/1600-0498.12306)]
58. Bautista JR, Zhang Y, Gwizdzka J. Healthcare professionals' acts of correcting health misinformation on social media. *Int J Med Inform* 2021 Apr;148:104375. [doi: [10.1016/j.ijmedinf.2021.104375](https://doi.org/10.1016/j.ijmedinf.2021.104375)] [Medline: [33461008](https://pubmed.ncbi.nlm.nih.gov/33461008/)]

## Abbreviations

**ALS:** amyotrophic lateral sclerosis

**CDC:** Centers for Disease Control and Prevention

**CMIS:** comprehensive model of information seeking

**CSM:** Common-Sense Model of Self-Regulation  
**HBM:** Health Belief Model  
**HISB:** health information-seeking behavior  
**IDK:** I do not know  
**MSE:** mean-square error

*Edited by A Mavragani; submitted 04.05.22; peer-reviewed by B Bagarić, Y Zhang; comments to author 20.07.22; revised version received 05.08.22; accepted 22.08.22; published 28.09.22.*

*Please cite as:*

*Korshakova E, Marsh JK, Kleinberg S*

*Health Information Sourcing and Health Knowledge Quality: Repeated Cross-sectional Survey*

*JMIR Form Res 2022;6(9):e39274*

URL: <https://formative.jmir.org/2022/9/e39274>

doi: [10.2196/39274](https://doi.org/10.2196/39274)

PMID: [35998198](https://pubmed.ncbi.nlm.nih.gov/35998198/)

©Elena Korshakova, Jessecae K Marsh, Samantha Kleinberg. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 28.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.

---

Publisher:  
JMIR Publications  
130 Queens Quay East.  
Toronto, ON, M5A 3Y5  
Phone: (+1) 416-583-2040  
Email: [support@jmir.org](mailto:support@jmir.org)

---

<https://www.jmirpublications.com/>