## Contents

### Original Papers

#### Comparing Self-Reported Running Distance and Pace With a Commercial Fitness Watch Data: Reliability Study (e39211)
Garrett Bullock, Joanne Stocks, Benjamin Feakins, Zahra Alizadeh, Amelia Arundale, Stefan Kluzek. 2

#### Prevalence of Body Dysmorphic Disorder in the Spanish Population: Cross-Sectional Web-Based Questionnaire Study (e46515)
Álvaro Loewen, Hilario Blasco-Fontecilla, Chao Li, Marcos Bella-Fernández, Belén Ruiz-Antorán. 11

#### Methodological Insights on Recruitment and Retention From a Remote Randomized Controlled Trial Examining the Effectiveness of an Alcohol Reduction App: Descriptive Analysis Study (e51839)
Melissa Oldham, Larisa Dinu, Gemma Loebenberg, Matt Field, Matthew Hickman, Susan Michie, Jamie Brown, Claire Garnett. 26

#### Remote Delivery of the Cuidándome Telehealth Intervention for Self-Management of Depression and Anxiety Among Latina Immigrant Women: Randomized Controlled Trial (e52969)
Carmen Alvarez, Subhash Aryal, Elizabeth Vrany, Maria Sanchez R, Rosalphie Quiles, Lia Escobar-Acosta, Felicia Hill-Briggs. 39

#### Feasibility and Acceptability of a Mobile Technology Intervention to Support Postabortion Care After Surgical Abortion (the FACTS Study Phase 3): Mixed Methods Prospective Pilot Study (e46284)
Roopan Gill, Gina Ogilvie, Wendy Norman, Brian Fitzsimmons, Cianna Maher, Regina Renner. 55

#### Machine Learning for Early Prediction of Major Adverse Cardiovascular Events After First Percutaneous Coronary Intervention in Patients With Acute Myocardial Infarction: Retrospective Cohort Study (e48487)
Pin Zhang, Lei Wu, Ting-Ting Zou, ZiXuan Zou, JiaXin Tu, Ren Gong, Jie Kuang. 69

#### Differences in Psychological Inflexibility Among Men With Erectile Dysfunction Younger and Older Than 40 Years: Web-Based Cross-Sectional Study (e45998)
Junichi Saito, Hiroaki Kumano, Mohammad Ghazizadeh, Chigusa Shimokawa, Hideki Tanemura. 80

#### Determining Distinct Suicide Attempts From Recurrent Electronic Health Record Codes: Classification Study (e46364)
Kate Bentley, Emily Madsen, Eugene Song, Yu Zhou, Victor Castro, Hyunjoon Lee, Younga Lee, Jordan Smoller. 88
Comparing Self-Reported Running Distance and Pace With a Commercial Fitness Watch Data: Reliability Study

Garrett Bullock1*, DPhil, DPT; Joanne Stocks2*, PhD; Benjamin Feakins3, DPhil; Zahra Alizadeh4, MD; Amelia Arundale5, PhD, DPT; Stefan Kluzek2, MD, DPhil

1Wake Forest School of Medicine, Winston-Salem, NC, United States
2University of Nottingham, Nottingham, United Kingdom
3University of Oxford, Oxford, United Kingdom
4Department of Sports Science and Exercise Medicine, Imam Khomeini Hospital Complex, Tehran University of Medical Sciences, Tehran, Iran
5Red Bull Athlete Performance Center, Thalgua, Austria
*these authors contributed equally

Corresponding Author:
Garrett Bullock, DPhil, DPT
Wake Forest School of Medicine
475 Vine St
Winston-Salem, NC, 27411
United States
Phone: 1 3367144264
Email: garrettbullock@gmail.com

Abstract

Background: There is substantial evidence exploring the reliability of running distance self-reporting and GPS wearable technology, but there are currently no studies investigating the reliability of participant self-reporting in comparison to GPS wearable technology. There is also a critical sports science and medical research gap due to a paucity of reliability studies assessing self-reported running pace.

Objective: The purpose of this study was to assess the reliability of weekly self-reported running distance and pace compared to a commercial GPS fitness watch, stratified by sex and age. These data will give clinicians and sports researchers insights into the reliability of runners’ self-reported pace, which may improve training designs and rehabilitation prescriptions.

Methods: A prospective study of recreational runners was performed. Weekly running distance and average running pace were captured through self-report and a fitness watch. Baseline characteristics collected included age and sex. Intraclass correlational coefficients were calculated for weekly running distance and running pace for self-report and watch data. Bland-Altman plots assessed any systemic measurement error. Analyses were then stratified by sex and age.

Results: Younger runners reported improved weekly distance reliability (median 0.93, IQR 0.92-0.94). All ages demonstrated similar running pace reliability. Results exhibited no discernable systematic bias.

Conclusions: Weekly self-report demonstrated good reliability for running distance and moderate reliability for running pace in comparison to the watch data. Similar reliability was observed for male and female participants. Younger runners demonstrated improved running distance reliability, but all age groups exhibited similar pace reliability. Running pace potentially should be monitored through technological means to increase precision.

(JMIR Form Res 2024;8:e39211) doi:10.2196/39211

KEYWORDS

GPS; Garmin; training load; running; exercise; fitness; wearables; running; running distance; pace; pace distance

Introduction

Physical activity is an essential component of a healthy lifestyle [1]. There is a substantial body of evidence highlighting the physical, social, and psychological health benefits of regular physical activity [1-3]. Sustainable physical activity interventions are needed, given that 31% of the global population is sedentary [4]. The World Health Organization’s physical activity action plan [5] identifies sport as an underused yet significant contributor to physical activity.
One widely popular sport globally is running [6]. Over the past 40 years, running has become one of the most popular physical leisure activities [7,8]. An estimated 50 million people in Europe participate in running as a way to stay healthy [9]. Due to high running participation prevalence [9], researchers have attempted to quantify running habits and training load, most notably through self-report [10]. Running load or workload is the distance run in 1 session. A training session is 1 running bout. Running speed is the intensity at which one runs for 1 running session [9,10]. However, there are potential inaccuracies from over–self-reporting due to recall bias [11] and social desirability of higher levels of physical activity [12], with potential differences by sex and age groups [13]. Further, the reliability of self-reported running pace has not been investigated, which is an important factor in quantifying running training intensity [14]. Due to these issues, research has investigated the reliability of wearables in quantifying running load [15]. Wearables, such as accelerometers, have demonstrated excellent reliability in assessing gait patterns, acceleration, and velocity [15].

Although wearable accelerometers are ubiquitously used in the general population [16] and are reliably used in research to measure physical activity levels [17], they are rarely used by running populations to track running load and training [18]. Runners opt for wearable GPS watches to track running training [19], with up to 90% of regular runners using some form of GPS monitoring when running [18]. GPS wearable technology quantifies running workload and speed [20]. A systematic review determined that there is excellent reliability for step counting and moderate validity for energy expenditure and distance run [21]. The most popular GPS wearable technology used by runners is the Garmin watch, as indicated by a previous survey where 44% of GPS and sports watches were Garmin, compared to 27% for Polar and 7% for Nike watches [22].

There has been previous related work in evaluating the reliability of running self-reports in large samples. In a sample of 92 endurance runners, followed for a 52-week (ie, 1 year) period, 93% of the runners participated in the entire follow-up period [10]. In a study of 53 running participants over 18 weeks, the response rate was 73% over the reporting period [23]. Another study surveyed 228 runners for at least 6 months, with a 2.2% attrition rate [24].

Methods

Study Design

A prospective cohort study was conducted using a mobile-based app. Participants accessed a dynamic digital consent form through the app or the recruitment website during the spring of 2021 over a 4-month period. During consent, participants could select different levels of study engagement. All levels of engagement involved the following: (1) an acknowledgement, understanding, and consent to participate in the study; (2) a baseline questionnaire collecting information on demographics, previous and current injury and illness history, footwear and foot posture, knee symptoms, lifestyle, and previous year’s training load; and (3) weekly reports on training load and each participant’s perceptions of cardiorespiratory symptoms, mood, and incidence of illness and injury in the last week. More advanced participation involved connecting participants’ Garmin Connect (Garmin Ltd) data, which included sharing data on running distance, running speed, and heart rate during each training or racing session. Participants added their Garmin Connect information at study recruitment. Garmin data were then automatically uploaded every week when the participant was within the study. Once the participant reaches the end of the study data collection or voluntarily leaves the study, the Garmin data collection link is terminated, ending data upload (Figure 1). Participants could opt out of the study at any time.
**Ethical Considerations**
This study received a favorable ethical review from the University of Nottingham (FMHS 113-1120). All methods were performed following the relevant guidelines and regulations of the Declaration of Helsinki. Before study inclusion, all participants were detailed about the risks and benefits of participation. All participants provided informed consent to participate.

**Population and Recruitment Strategy**
This study aimed to examine recreational runners. The inclusion criteria of this subgroup of the larger “Running Through” [26] study consisted of the following: (1) age ≥18 years; (2) performing running activities; and (3) connecting their Garmin Connect data to the weekly reports. Exclusion criteria consisted of individuals meeting the following conditions: (1) not able or willing to use the internet regularly; (2) diagnosed with an immunocompromised disorder; (3) diagnosed with memory impairment; (4) diagnosed with a neurodegenerative disorder; (5) diagnosed with inflammatory osteoarthritis; and (6) undergone trunk or lower extremity orthopedic surgery in the last 6 months. The larger “Running Through” study consisted of both Garmin and self-report data. Participants were recruited through email, the study website, social media, and word of mouth. Participants resided in the United Kingdom or Europe. All recruitment was performed in English. Participants did not receive compensation for participating in this study. Watch ownership was not known by the research team.

**Weekly Survey**
Participants were sent an encrypted text message or email weekly to report weekly running distance, pace, and incidence of illness and injury [10]. Garmin Connect data also monitored running distance and pace. Garmin monitoring has demonstrated excellent reliability and validity [20].

**Data Storage**
The University of Nottingham’s secure server hosted the research survey tool through the RedCap (Research Electronic Data Capture) service [27]. Data were queried from the secure database using a unique randomized and encrypted identification number.

**Data Reduction**
Watch data were downloaded to an encrypted SQL database using Garmin Connect software (Garmin Ltd). For convenience, these data were combined with the RedCap survey data and tables containing key variables that could be used to link these data. Custom functions were written in R using the DBI and MariaDB packages to interface with the database. The rjson and bit64 packages were additionally used to facilitate the extraction of JavaScript Object Notation format activity data and provide necessary extensions to R’s base data classes. Once data were downloaded, they were aggregated, cleaned, and checked for quality assurance. Data checks were performed through automation and manually.

**Statistical Analyses**
Participant statistics were described using mean (SD) or median (IQR) for continuous variables and frequencies (percentages)
for categorical variables. Overall running exposure was calculated in person kilometers.

To assess reliability, intraclass correlational coefficients (ICC_{2,1}) were calculated for weekly running distance and running pace between self-reports and weekly reports generated by the Garmin Connect data. Reliability was rated as poor (<0.50), moderate (0.50-0.75), good (0.75-0.90), and excellent (>0.90) [28]. Correlation and Bland-Altman plots were also calculated to assess any systematic measurement error. Analyses were then stratified by sex and age strata (18-40, 41-60, and ≥61 years). All analyses were performed in R 4.1.2 (R Foundation for Statistical Computing) [29], with the psych package for ICC calculations and BlandAltmanLeh for Bland-Altman plots.

### Table 1. Participant descriptive statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall (n=475)</th>
<th>Female participants (n=248)</th>
<th>Male participants (n=227)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>49.5 (12.2)</td>
<td>51.0 (13.1)</td>
<td>47.8 (10.9)</td>
</tr>
<tr>
<td>BMI (kg/m^2), mean (SD)</td>
<td>24.0 (3.7)</td>
<td>24.2 (3.7)</td>
<td>23.8 (3.8)</td>
</tr>
<tr>
<td>Number of weeks followed, median (IQR)</td>
<td>17 (11-23)</td>
<td>15 (10-20)</td>
<td>17 (11-24)</td>
</tr>
</tbody>
</table>

### Smoking, n (%)

| Current smoker | 10 (2) | 6 (2) | 4 (1) |
| Ex-smoker      | 62 (13) | 30 (13) | 32 (13) |
| Cigarettes per day, median (IQR)      | 9 (3-14) | 7 (5-10) | 30 (23-37) |
| Years smoked, median (IQR)            | 22 (9-35) | 14 (2-26) | 25 (18-31) |
| Diabetes, n (%)                        | 6 (1) | 1 (1) | 5 (3) |
| Heart disease, n (%)                   | 5 (1) | 2 (1) | 3 (1) |
| Cancer, n (%)                           | 14 (3) | 10 (4) | 4 (2) |
| Asthma, n (%)                           | 62 (13) | 40 (16) | 32 (11) |
| Hay fever (pollen allergies), n (%)    | 180 (38) | 99 (40) | 81 (35) |
| Days of running per week, mean (SD)    | 2 (1) | 2 (1) | 2 (1) |

### Table 2. Weekly running descriptive characteristics using Garmin watch data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall (n=475)</th>
<th>Female participants (n=248)</th>
<th>Male participants (n=227)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time running (min), median (IQR)</td>
<td>144 (48-238)</td>
<td>139 (43-226)</td>
<td>153 (45-230)</td>
</tr>
<tr>
<td>Calories burned (kcal), median (IQR)</td>
<td>1450 (547-2353)</td>
<td>1212 (461-1963)</td>
<td>1755 (671-2840)</td>
</tr>
<tr>
<td>Kilometers, median (IQR)</td>
<td>25.9 (4.7-41.8)</td>
<td>22.7 (4.5-40.9)</td>
<td>29.6 (4.9-40.6)</td>
</tr>
<tr>
<td>Running pace (min/km), median (IQR)</td>
<td>6.1 (5.2-7.0)</td>
<td>6.7 (5.7-7.7)</td>
<td>5.7 (4.9-6.4)</td>
</tr>
<tr>
<td>Average heart rate (bpm(^a)), mean (SD)</td>
<td>130 (26)</td>
<td>127 (26)</td>
<td>130 (24)</td>
</tr>
<tr>
<td>Maximum heart rate (bpm(^a)), mean (SD)</td>
<td>163 (24)</td>
<td>162 (25)</td>
<td>163 (24)</td>
</tr>
</tbody>
</table>

\(^a\)Bpm: beats per minute.

### Results

A total of 485 participants linked their Garmin Connect data to the study, with 475 participants included for a total of 3602 participant weeks. Of these, 3 participants were excluded due to lack of follow-up, and another 7 did not run during the collection period (Table 1; the flowchart is available in Multimedia Appendix 1). Participants self-reported running a weekly median of 26.2 (IQR 12.8-39.7) km at a median pace of 6.0 (IQR 5.4-6.7) min/km compared to 25.9 (IQR 4.7-41.8) km running distance at a median pace of 6.1 (IQR 5.2-7.0) min/km (Table 2) recorded by the Garmin watch.

### Reliability

Weekly distance and pace reliability were rated as good and moderate, respectively, for both sexes and for runners aged 41-60 and ≥61 years. Furthermore, weekly distance reliability was rated as excellent and moderate in runners aged 18-40 years. All results exhibited no discernable systematic bias (Figure 2; Table 3; Multimedia Appendix 1).
Figure 2. Correlation and Bland-Altman Plots of the Reliability of Self-Report and Garmin Connect Weekly Running Distance and Running Pace. A. Weekly Running Distance (km) B. Weekly Running Pace (min/km).
Table 3. Reliability of weekly self-report and Garmin watch data for running distance and pace.

<table>
<thead>
<tr>
<th>Group</th>
<th>Self-report</th>
<th>Garmin watch</th>
<th>ICCa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weekly distance run (km), median (IQR)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall (n=475)</td>
<td>26.2 (12.8, 39.7)</td>
<td>25.9 (4.7-41.8)</td>
<td>0.88 (0.87-0.89)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (n=248)</td>
<td>24.8 (10.0-39.7)</td>
<td>22.7 (4.5-40.9)</td>
<td>0.86 (0.85-0.87)</td>
</tr>
<tr>
<td>Male (n=227)</td>
<td>29.0 (16.9-41.1)</td>
<td>29.6 (4.9-40.6)</td>
<td>0.89 (0.88-0.90)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-40 (n=113)</td>
<td>27.0 (10.9-43.0)</td>
<td>26.8 (7.2-44.6)</td>
<td>0.93 (0.92-0.94)</td>
</tr>
<tr>
<td>41-60 (n=262)</td>
<td>27.0 (16.0-38.0)</td>
<td>25.9 (11.5-41.2)</td>
<td>0.87 (0.85-0.88)</td>
</tr>
<tr>
<td>≥61 (n=100)</td>
<td>25.0 (13.8-36.2)</td>
<td>24.5 (9.7-39.3)</td>
<td>0.83 (0.80-0.85)</td>
</tr>
<tr>
<td><strong>Average weekly running pace (min/km), mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall (n=475)</td>
<td>6.0 (1.1)</td>
<td>6.1 (2.0)</td>
<td>0.72 (0.69-0.75)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (n=248)</td>
<td>6.4 (1.2)</td>
<td>6.7 (1.9)</td>
<td>0.67 (0.62-0.72)</td>
</tr>
<tr>
<td>Male (n=227)</td>
<td>5.7 (0.9)</td>
<td>5.7 (2.1)</td>
<td>0.68 (0.65-0.71)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-40 years (n=113)</td>
<td>5.8 (1.1)</td>
<td>5.7 (2.5)</td>
<td>0.69 (0.65-0.73)</td>
</tr>
<tr>
<td>41-60 years (n=262)</td>
<td>6.0 (1.0)</td>
<td>6.1 (2.5)</td>
<td>0.70 (0.66-0.73)</td>
</tr>
<tr>
<td>≥61 years (n=100)</td>
<td>6.2 (1.2)</td>
<td>6.5 (2.5)</td>
<td>0.74 (0.68-0.78)</td>
</tr>
</tbody>
</table>

aICC: intraclass correlation coefficient.

**Discussion**

**Principal Findings**

The overall findings of this study indicate that the weekly self-report of running distance by runners wearing a Garmin watch demonstrated good reliability compared to the Garmin watch data. Distance reliability was similar between female and male participants and across age strata, except for participants aged 18-40 years, who demonstrated excellent reliability. Weekly self-report of running pace demonstrated moderate reliability compared to Garmin watch data, with similar reliability observed between sex and age strata. There were no discernable patterns or systematic biases concerning self-reported running distance or pace.

**Comparison to Previous Work**

Self-reported running distance exhibited good reliability compared to Garmin data. The reliability is higher in this study compared to a previous study on physical activity (ICC 0.67-0.81) [30]. However, the previous study examined multiple countries and recorded all physical activity beyond running. The homogenous country sample and the focus on running in our study may affect the comparison of these results [31]. Younger adults (aged 18-40 years) demonstrated increased running distance reliability reporting compared to the older age strata (aged 41-60 and ≥61 years). This is comparable with previous research, in which younger adults displayed improved self-report reliability [30]. Younger adults may have a greater aptitude to monitor their running through technology [32]. However, this is only speculative, and further research is required.

Self-reported running pace demonstrated moderate reliability compared to Garmin Connect data. There are currently no studies investigating the reliability of self-reported running pace. However, recreational runners usually train at one pace, with little change at different distances [33]. The moderate reliability observed in this sample may be due to these runners reporting their perceived running pace, with little fluctuation between sessions or weeks. However, specific variances may have occurred in the actual running pace, decreasing the reliability of these data. Previous literature has suggested that instant feedback through the use of heart rate or step cadence can increase a recreational runner’s ability to self-report running pace [33]. However, further work is needed to investigate the efficacy of this approach.

These findings necessitate future research. Participants were recreational runners, and most of them were older than 40 years. Future work is needed to assess the reliability of self-report in comparison to GPS monitoring data in elite runners of all ages and younger populations across different skill or competition levels. All runners in this study already owned a Garmin watch before the study enrollment. Understanding how self-reporting changes among new GPS activity monitor users is needed. Running pace demonstrated moderate reliability in this recreational runner sample. Future research is required to investigate the effectiveness of running pace training on self-report reliability.
Limitations
As with all studies, there are limitations to this study. First, there is the risk of recall bias due to the weekly intervals for self-reporting, which decreases the precision of these findings. Participants may not have worn or activated their Garmin watch for specific runs, decreasing the reliability of these data. GPS monitoring devices are expensive, causing a high barrier to entry. Such a barrier may add selection bias to this reliability study, decreasing the generalizability of these results to all running populations. Further, the sample in this study comprised recreational runners; therefore, the results are not generalizable to elite runners or populations that solely engage in walking for exercise. Finally, participants used different versions of the Garmin watch. As different watch versions may exhibit different reliability, there is a potential for decreased data precision.

Practical Applications
Physical activity monitors have effectively enhanced physical activity levels by providing user feedback, facilitating behavior change—following prescribed training—and preventing injuries [34]. The good to excellent reliability of self-reported weekly running distance observed in this cohort of recreational runners across all adult age groups supports previous research indicating that runners can effectively report running loads [30]. These findings strengthen the notion that self-report can be used to reliably monitor runners as they begin or maintain an exercise regimen or return to running following an injury. However, the moderate reliability exhibited for running pace suggests that recreational runners of all ages are not as adept at monitoring their running pace. Incorporating technological monitoring for running pace may be pertinent to maintain prescribed running paces either for specific training regimens (ie, preparing for a race) or when returning to running following an injury.

Conclusions
Weekly self-report demonstrated good reliability for running distance compared to the Garmin watch data, with similar reliability between sex and age groups. However, the weekly self-report demonstrated only moderate reliability for running pace compared to the Garmin data, with similar reliability between sex and age groups. Sports researchers and scientists can use weekly self-reported running distance in conjunction with Garmin data when quantifying weekly training load. However, caution should be exercised when relying on self-reported running pace to evaluate running intensity in recreational runners. Running pace potentially should be monitored through technological means to increase precision.

Data Availability
Data and corresponding codes are available within the Open Science Framework [35].

Acknowledgments
This study was funded through Technopolis Consulting Group Belgium (TGB).

Authors' Contributions
GB, JS, BF, and SK conceived the study idea. JS, ZA, and SK collected the data. GS, BF, JS, and SK extracted and analyzed the data. GB, JS, BF, AA, and SK wrote the original manuscript. GB, JS, BF, ZA, AA, and SK edited the manuscript. GB, JS, BF, ZA, AA, and SK approved the final draft of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Study flowchart.
[DOCX File, 25 KB - formative_v8i1e39211_app1.docx ]

References


24. Running through study. University of Nottingham. URL: https://www.nottingham.ac.uk/research/groups/crsm/running-through-study.aspx?--text=The%20Running%20Through%20research%20team.load%2C%2ointensity%20or%2oinfection%20recovery [accessed 2021-01-01]


26. Running through study. University of Nottingham. URL: https://www.nottingham.ac.uk/research/groups/crsm/running-through-study.aspx?--text=The%20Running%20Through%20research%20team.load%2C%2ointensity%20or%2oinfection%20recovery [accessed 2021-01-01]

35. Open Science Framework. URL: https://osf.io/jgfau/ [accessed 2023-12-15]

Abbreviations

ACC: intraclass correlational coefficient
RedCAP: Research Electronic Data Capture
Prevalence of Body Dysmorphic Disorder in the Spanish Population: Cross-Sectional Web-Based Questionnaire Study

Álvaro Loewen¹, MD; Hilario Blasco-Fontecilla¹,²,³,⁴, MD, PhD; Chao Li¹, MD; Marcos Bella-Fernández⁵,⁶, MSc; Belén Ruiz-Antorán⁷, MD, PhD

¹Facultad de Medicina, Universidad Autónoma de Madrid, Madrid, Spain
²Servicio de Psiquiatría Infanto-juvenil, Hospital Universitario Puerta de Hierro Majadahonda, Majadahonda, Spain
³Centro de Investigación Biomédica en Red - Salud Mental, Madrid, Spain
⁴Facultad de Ciencias de la Salud y Centro Médico, Universidad Internacional de La Rioja, Madrid, Spain
⁵Facultad de Psicología, Universidad Autónoma de Madrid, Madrid, Spain
⁶Departamento de Psicología, Universidad Pontificia de Comillas, Madrid, Spain
⁷Servicio de Farmacología Clínica, Hospital Universitario Puerta de Hierro Majadahonda, Majadahonda, Spain

Corresponding Author:
Hilario Blasco-Fontecilla, MD, PhD
Servicio de Psiquiatría Infanto-juvenil
Hospital Universitario Puerta de Hierro Majadahonda
Instituto de Investigación Sanitaria Puerta de Hierro - Segovia de Arana
C/ Manuel de Falla, 1
Majadahonda, 28222
Spain
Phone: 34 911911690
Email: hmblasco@yahoo.es

Abstract

Background: Body dysmorphic disorder (BDD) is defined as excessive concern with mild or nonexistent defects in personal physical appearance, which are not perceived by others. The worldwide prevalence of BDD ranges between 0.5% and 3.2%, with no differences across genders. The mean age of onset of BDD is 16.9 years. BDD is typically associated with young age, psychiatric disorders, and dermatological procedures. Patients with BDD typically display poorer mental health status than patients diagnosed with other mental disorders.

Objective: The aim of this study was to estimate the prevalence of BDD in Spain and to identify the variables associated with BDD.

Methods: We performed a cross-sectional descriptive study by collecting data through an anonymous web-based survey targeting the Spanish population aged 18 years or older. The measures in this study were (1) sociodemographic variables, (2) variables associated with dermatological and psychiatric disorders and cosmetic procedures, (3) scales measuring quality of life (12-item Short Form health survey, version 2) and (4) BDD (BDD Questionnaire). Statistical analysis was performed with SPSS software version 21. P values less than .05 were considered significant.

Results: Of the 2091 participants who took the survey, 322 (15.2%) met the criteria of having BDD. The mean age of the participants with BDD was 23.5 (SD 9.6) years. In terms of BDD prevalence, women accounted for 19.9% (284/1421), men accounted for 5.2% (34/653), and students accounted for 25.2% (263/1043). Approximately 46.6% (150/322) of the participants with BDD reported a history of psychiatric comorbidities, including anxiety disorders, depressive disorders, and eating disorders. BDD was significantly associated with female gender, younger age (18-24 years), students, monthly income of less than €500 (€1=US $1.11), and the presence of dermatological and some psychiatric disorders such as depression, anxiety, and eating disorders (P<.05). The number of body parts of concern in participants with BDD was significantly higher than that in those without BDD (4.6 vs 2.2, respectively; P<.001). Regarding the body parts of concern, body fat was the most common concern for both groups with BDD and without BDD, followed by thighs, face, hips, and skin in the BDD group and thighs, teeth, and hair in the non-BDD group. Participants with BDD showed a significantly poorer self-perception of their mental health, irrespective of the presence of any mental disorder (P<.001).
Conclusions: Our findings showed that the prevalence of BDD in Spain was higher than expected. Further, BDD is frequently associated with other psychiatric disorders, particularly depressive disorder, anxiety disorder, and eating disorder. Participants with BDD had a poorer perception of quality of life associated with mental but not physical health problems. Finally, the perception of quality of mental health life in participants with BDD was independent of diagnosis of any mental disorder.

(JMIR Form Res 2024;8:e46515) doi: 10.2196/46515

KEYWORDS
body dysmorphic disorder; prevalence; adults; Spain; comorbidities; mental health; depression; anxiety; OCD; obsessive-compulsive disorder

Introduction

Body dysmorphic disorder (BDD) is a common psychiatric disorder affecting 0.5%-3.2% of the general population worldwide [1]. A multicentric study in Spain showed that the prevalence of BDD was higher among patients with acne (10.6%) [2]. The prevalence of BDD across genders is debatable. One study showed similar prevalence of BDD across both genders (49% females and 51% males) [3]. Another study reported higher prevalence among females (68.5%) [4]. The average age of onset of BDD is 16.9 years [5]. One study reported an inverse relationship between the prevalence of BDD and age: 78.6% of participants with BDD were aged 18-27 years, 14.3% were aged 28-37 years, and 7.1% were aged 38 years or older [6]. Furthermore, BDD is closely associated with other mental disorders. A recent systematic review [1] showed that the highest prevalence of BDD was among psychiatric inpatients (5.8%-37.78%). The psychiatric disorders most frequently associated with BDD are depressive disorder (47%-56.3%), borderline personality disorder (54.3%), and eating disorders (12%-45%), whereas obsessive-compulsive disorder (3%-15.3%) and schizophrenia are less closely associated with BDD. Participants who had cosmetic procedures (2.9%-57%) slightly overlay with BDD prevalence in the general population. The prevalence of BDD in both psychiatric outpatients (0.3%-2%) and students (1.3%-5.8%) partially overlaps with that observed in the general population. Among dermatologic patients, the prevalence of BDD was reported to be 2.1%-36% [1].

The etiology of BDD is multifactorial, encompassing biological, psychological, and sociocultural factors. BDD has been associated with parental rejection, as well as physical, emotional, or sexual abuse during adolescence [7]. Studies have shown a possible genetic association in first-degree relatives, with affected patients being up to 3-8 times more likely to develop BDD than the general population [7]. Shy, anxious, and perfectionistic individuals may also have a greater predisposition to develop this disorder [8].

The most important symptom of BDD is distortion of body perception, which leads to low self-esteem, anxiety, depression, social isolation, and obsessive-compulsive behaviors [9]. The clinical profile of BDD is characterized by repetitive actions such as constant checking in mirrors, applying excessive makeup to cover defects, dermatillomania, comparing one's appearance with that of others, and excessive exercise, taking up an average of 3-8 hours daily [10]. Patients with BDD are usually preoccupied with 5-7 different parts of their body [11], the most common being skin (53.8%), nose (38.5%), and hair (34.6%). The other body parts that are of frequent concern are weight and muscle mass (30.8%), face (30.8%), chest and trunk (19.2%), and teeth (19.2%) [1]. The mental and physical health status perceived by patients with BDD are lower than that perceived by the general population [12]. Given that the primary concern of patients with BDD revolves around their external appearance, it is common for people with BDD to predominantly seek dermatological and cosmetic procedures over seeking professional help for the treatment of their underlying psychiatric pathology [13]. Moreover, patients diagnosed with BDD often do not seek help for various reasons: they feel ashamed or lack insight [14]. Furthermore, BDD is likely to be underdiagnosed, given the large number of barriers to diagnosis such as the absence of appropriate tools, lack of information and awareness among health care professionals, and professionals' refusal to lose a patient or inability to diagnose it properly [15]. In addition, not all health care professionals are familiar with this disorder [9]. This leads to poor identification in psychiatric and cosmetic settings where BDD cases are notoriously prevalent [14]. If professionals do not perform a detailed anamnesis, it is difficult for patients to disclose their concerns, given the shyness underlying this disorder [16]. To overcome this shortcoming, clinicians may use the BDD Questionnaire (BDDQ), a validated diagnostic tool for BDD, with sensitivity of 100% [17-19], specificity of 89%-93% for psychiatric inpatients [18], and 93% for dermatologic patients [19]. The BDDQ is a brief questionnaire that assesses different items of the patient's body perception [17]. Finally, treatment is based on pharmacological and nonpharmacological measures. The former includes the use of fluoxetine, a serotonin reuptake inhibitor antidepressant. The latter is based on psychotherapy—most notably cognitive behavioral therapy [20]. Randomized clinical trials have shown a rate of 50%-80% improvement in patients as well as a lower relapse rate following pharmacological treatment [21].

In summary, BDD is a poorly studied and underdiagnosed psychiatric disorder. This may be because BDD is not perceived as a disorder by aesthetic practitioners, who may think that they are merely offering a “service” [22-24]. The main objective of our study was to estimate the prevalence of BDD in Spain. Additionally, we explored the association of BDD with sociodemographic variables, presence of dermatological or psychiatric disorders and cosmetic procedures, and quality of life.
Methods

Design and Scope of This Study
A cross-sectional descriptive study was conducted. The general population aged 18 years and older in Spain was invited to access the study protocol via a link to Google Forms. The security and lawful use of personal data collected on the website is guaranteed by accepting the data privacy policy included in the survey. The measures in the protocol consisted of (1) sociodemographic variables, (2) variables associated with dermatological and psychiatric disorders and cosmetic procedures, and (3) perception of health and quality of life, measured using the 12-item Short Form version 2 (SF-12v2) health survey, and (4) a validated scale for the diagnosis of BDD, using the BDD questionnaire screening test [17].

Study Sample Population
Participants 18 years and older from the general population residing and registered in Spain at the time the survey was performed and who voluntarily completed the study’s web-based questionnaire were included. The sample size was calculated using Epidat 4.0 (Dirección Xeral de Saúde Pública da Consellería de Sanidade da Xunta de Galicia) based on the following estimate: population size, 40,000,000; expected proportion, 5%; accuracy, 1%; confidence level, 95%; and design effect, 1.0. The minimum total number of responses required for 1% precision with 95% confidence level was 1825 participants.

Variables
The following variables were collected: sociodemographic variables such as gender (female, male, or other, ie, participants who do not identify themselves as male or female), year of birth, region of residence, race, educational level, employment status, and range of monthly income. Variables related to other comorbidities associated with BDD were comorbidity with dermatologic and psychiatric diseases and cosmetic procedures. Regarding the year of birth, the participants were classified into 4 age groups: 18-24 years, 25-44 years, 45-64 years, and 65 years or older. This classification was based on previous studies related to mental health and the use of these services according to age group [25].

Quality of Life Assessment
Data on quality of life and perception of their state of health were collected through the SF-12v2 health survey, which is validated as a psychometric instrument for numerous diseases and conditions, including mental illness. It assesses the participant’s physical and mental state through 8 health domains: 4 related to physical health, that is, general health, physical function, role-physical, and bodily pain; and 4 related to mental health, that is, vitality, social function, role-emotional, and mental health [26].

BDD Assessment
BDD assessment was performed using the BDDQ [17]. A version adapted and validated in Spanish was used [27]. The BDDQ is a brief (7-item) self-report measure derived from the Diagnostic and Statistical Manual of Mental Disorders, fourth edition diagnostic criteria. The first 6 items require a dichotomous answer (yes/no) and the last one is multiple choice. The test will be positive if the patient answers “yes” to questions 1 and 2, “yes” to question 3, 4, 5, or 6, and if in question 7, the time indicated is more than 1 hour per day [17].

Data Handling
The Google Forms platform was used for the web-based survey. The questionnaire was distributed telematically, both through a link via mobile phone and a printed QR code. The data obtained were extracted and sorted in Microsoft Excel. Subsequently, we used SPSS to create the database and perform the corresponding analyses. A license was obtained for the SF-12v2 health survey, which together with the use of the QualityMetric program provided, allowed for its correct interpretation. The database was generated in an anonymized form guaranteeing exclusive access by the principal investigator, thereby allowing respect, privacy, and confidentiality of the data.

Statistical Analysis
Statistical analysis was performed with SPSS software (version 21; IBM Corp). All statistical analyses were performed at .05 level of significance. A descriptive study was conducted for all the variables included in this study. Quantitative variables were expressed as mean and standard deviation (SD). Qualitative variables were expressed as absolute value (n) and percentage with an estimated 95% CI. Comparison of means was performed using 2-sided Student t test or Mann-Whitney U test as appropriate after checking normality with the Kolmogorov-Smirnov test. The association of the qualitative variables was estimated by means of the chi-square statistic. Multiple logistic regression models were used to determine the association of different variables with each other. A univariate analysis was performed where a significantly higher risk ratio for BDD diagnosis was obtained for some of the variables studied. The significant variables obtained in the univariate model were subsequently included in the multivariate analysis. Thereafter, given the heterogeneous conditions of the population, a subgroup analysis was performed with a multivariate model considering the gender and age variables.

Ethics Approval
This study was conducted in accordance with the requirements expressed in the Declaration of Helsinki 2013. Participants were invited to participate online by clicking on the link to the survey. Information about the purpose of the survey and its anonymous and voluntary nature was included in the survey header. Participants were identified by a numerical code in order to respect the confidentiality of the participants’ personal data. The automatic code is assigned directly by Google Forms at the time of download through a time stamp. This project was approved by the ethics committee of the Hospital Universitario Puerta de Hierro Majadahonda in Madrid (promoter protocol code PI 206/21, December 20, 2021).
Results

A total of 2091 participants were included in this study. The sociodemographic and clinical characteristics of the participants are described in Table 1.

The prevalence of BDD in the population assessed in this study was 15.2% (284/1421, 19.9% in females vs 34/653, 5.2% in males). Regarding age, the prevalence of BDD was higher in the youngest age group (18-24 years; 267/1091, 24.5%), followed by the 25-44 years age group (30/279, 10.8%) (Figure 1). The number of body parts of concern in participants with BDD was significantly higher than that in participants without BDD (4.6 vs 2.2, respectively; \( P < .001 \)). Regarding the body parts of concern, body fat was the most common concern in both groups, followed by thighs, face, hips, and skin in the BDD group and thighs, teeth, and hair in the non-BDD group (Figure 2). Regarding the sociodemographic characteristics of the BDD population (n=322), the majority were females (284/322, 88.2%), with a mean age of 23.5 (SD 9.6) years, and Caucasians (243/722, 75.5%). Approximately 81.7% (1533/2091) were students, and 76.6% (247/322) of them had a university education level. Among the participants diagnosed with BDD, 63.4% (204/322) had a history of dermatologic disease, the most frequent being acne (115/204, 56.5%) and dermatitis (99/204, 48.5%). Approximately 46.6% (150/322) of the participants with BDD reported a history of psychiatric comorbidities, and the most frequent were anxiety disorders (108/150, 72%), depressive disorders (84/150, 56%), eating disorders (72/150, 48%), and attention-deficit/hyperactivity disorder (18/150, 12%). Approximately 17.7% (57/322) of the population with BDD had previously undergone a cosmetic procedure, the most frequent being laser treatment for acne, blemishes, and other skin lesions (17/57, 29.8%); mesotherapy (9/57, 15.8%); and rhinoplasty (7/57, 12.3%). The factors related to BDD are reported in univariate and multivariate analyses in Table 2. BDD diagnosis was significantly associated with female gender, other genders, age 18-24 years, students, monthly income level of less than €500 (€1=US $1.11), and participants with dermatologic and psychiatric comorbidities (\( P < .001 \)). All these variables were included in the multivariate model, with gender (female and other), age, student occupation, depressive disorder, eating disorders, and anxiety disorders remaining at <.05 significance as diagnostic predictors of BDD.
Table 1. Sociodemographic and clinical characteristics of the participants.

<table>
<thead>
<tr>
<th></th>
<th>All (N=2091)</th>
<th>With BDD&lt;sup&gt;a&lt;/sup&gt; diagnosis (n=322)</th>
<th>Without BDD diagnosis (n=1769)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender (female), n (%)</strong></td>
<td>1421 (68)</td>
<td>284 (88.2)</td>
<td>1137 (64.3)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>37.7 (16.6)</td>
<td>23.5 (9.6)</td>
<td>35.5 (16.9)</td>
</tr>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>1091 (52.2)</td>
<td>267 (82.9)</td>
<td>824 (46.6)</td>
</tr>
<tr>
<td>25-44</td>
<td>279 (13.3)</td>
<td>30 (9.3)</td>
<td>249 (14.1)</td>
</tr>
<tr>
<td>45-64</td>
<td>654 (31.3)</td>
<td>23 (7.1)</td>
<td>631 (35.7)</td>
</tr>
<tr>
<td>&gt;64 years</td>
<td>66 (3.2)</td>
<td>2 (0.6)</td>
<td>64 (3.6)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African</td>
<td>11 (0.5)</td>
<td>3 (0.9)</td>
<td>8 (0.5)</td>
</tr>
<tr>
<td>American</td>
<td>6 (0.3)</td>
<td>0 (0)</td>
<td>6 (0.3)</td>
</tr>
<tr>
<td>Asian</td>
<td>26 (1.2)</td>
<td>7 (2.2)</td>
<td>19 (1.1)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>1533 (73.3)</td>
<td>243 (75.5)</td>
<td>1290 (72.9)</td>
</tr>
<tr>
<td>Latin</td>
<td>306 (14.6)</td>
<td>23 (7.1)</td>
<td>283 (16)</td>
</tr>
<tr>
<td>Other</td>
<td>209 (10)</td>
<td>46 (14.3)</td>
<td>163 (9.2)</td>
</tr>
<tr>
<td><strong>Educational level, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td>11 (0.5)</td>
<td>1 (0.3)</td>
<td>10 (0.6)</td>
</tr>
<tr>
<td>Middle school</td>
<td>8 (0.4)</td>
<td>0 (0)</td>
<td>8 (0.5)</td>
</tr>
<tr>
<td>Professional education</td>
<td>100 (4.8)</td>
<td>15 (4.7)</td>
<td>85 (4.8)</td>
</tr>
<tr>
<td>High school</td>
<td>265 (12.7)</td>
<td>59 (18.3)</td>
<td>206 (11.6)</td>
</tr>
<tr>
<td>University</td>
<td>1707 (81.6)</td>
<td>247 (76.7)</td>
<td>1460 (82.5)</td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>1043 (49.9)</td>
<td>263 (81.7)</td>
<td>780 (44.1)</td>
</tr>
<tr>
<td>Worker</td>
<td>865 (41.4)</td>
<td>50 (15.5)</td>
<td>815 (46.1)</td>
</tr>
<tr>
<td>Other</td>
<td>183 (8.8)</td>
<td>9 (2.8)</td>
<td>174 (9.8)</td>
</tr>
<tr>
<td><strong>Monthly income (€)&lt;sup&gt;b&lt;/sup&gt;, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;500</td>
<td>861 (41.2)</td>
<td>205 (63.7)</td>
<td>656 (37.1)</td>
</tr>
<tr>
<td>500-999</td>
<td>138 (6.6)</td>
<td>34 (10.6)</td>
<td>104 (5.9)</td>
</tr>
<tr>
<td>1000-1999</td>
<td>343 (16.4)</td>
<td>39 (12.1)</td>
<td>304 (17.2)</td>
</tr>
<tr>
<td>2000-2999</td>
<td>291 (13.9)</td>
<td>9 (2.8)</td>
<td>282 (15.9)</td>
</tr>
<tr>
<td>&gt;3000</td>
<td>358 (17.1)</td>
<td>11 (3.4)</td>
<td>347 (19.6)</td>
</tr>
<tr>
<td>Not known</td>
<td>100 (4.8)</td>
<td>24 (7.5)</td>
<td>76 (4.3)</td>
</tr>
<tr>
<td><strong>Dermatologic disease, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acne</td>
<td>1072 (51.3)</td>
<td>204 (63.4)</td>
<td>868 (49.1)</td>
</tr>
<tr>
<td>Atopic dermatitis</td>
<td>532 (49.6)</td>
<td>115 (56.5)</td>
<td>417 (48)</td>
</tr>
<tr>
<td>Other dermatitis</td>
<td>434 (40.5)</td>
<td>99 (48.5)</td>
<td>335 (38.6)</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>35 (3.3)</td>
<td>10 (4.9)</td>
<td>25 (2.9)</td>
</tr>
<tr>
<td>Rosacea</td>
<td>50 (4.7)</td>
<td>2 (1)</td>
<td>48 (5.5)</td>
</tr>
<tr>
<td>Urticaria</td>
<td>64 (6)</td>
<td>5 (2.5)</td>
<td>59 (6.8)</td>
</tr>
<tr>
<td>Pityriasis</td>
<td>15 (1.4)</td>
<td>1 (0.5)</td>
<td>14 (1.6)</td>
</tr>
<tr>
<td>Eczema</td>
<td>37 (3.5)</td>
<td>9 (4.4)</td>
<td>28 (3.2)</td>
</tr>
<tr>
<td>Skin infections</td>
<td>32 (3)</td>
<td>6 (2.9)</td>
<td>26 (3)</td>
</tr>
<tr>
<td></td>
<td>All (N=2091)</td>
<td>With BDD(^a) diagnosis (n=322)</td>
<td>Without BDD diagnosis (n=1769)</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------</td>
<td>----------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>11 (1)</td>
<td>0 (0)</td>
<td>11 (1.3)</td>
</tr>
<tr>
<td>Other</td>
<td>65 (6.1)</td>
<td>7 (3.4)</td>
<td>58 (6.7)</td>
</tr>
<tr>
<td><strong>Psychiatric disorder, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressive Disorder</td>
<td>493 (23.6)</td>
<td>150 (46.6)</td>
<td>343 (19.4)</td>
</tr>
<tr>
<td>Borderline personality disorder</td>
<td>244 (49.5)</td>
<td>84 (56)</td>
<td>160 (46.6)</td>
</tr>
<tr>
<td>Eating disorder</td>
<td>9 (1.8)</td>
<td>3 (2)</td>
<td>6 (1.7)</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>33 (6.7)</td>
<td>12 (8)</td>
<td>21 (6.1)</td>
</tr>
<tr>
<td>Attention-deficit/hyperactivity disorder</td>
<td>59 (12)</td>
<td>18 (12)</td>
<td>41 (12)</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>7 (1.4)</td>
<td>1 (0.7)</td>
<td>6 (1.7)</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td>292 (59.2)</td>
<td>108 (72)</td>
<td>184 (53.6)</td>
</tr>
<tr>
<td>Schizophrenia/psychosis</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>12 (2.4)</td>
<td>3 (2)</td>
<td>9 (2.6)</td>
</tr>
<tr>
<td>Other</td>
<td>18 (3.7)</td>
<td>3 (2)</td>
<td>15 (4.4)</td>
</tr>
<tr>
<td>History of plastic surgery procedures, n (%)</td>
<td>355 (17)</td>
<td>57 (17.7)</td>
<td>298 (16.8)</td>
</tr>
</tbody>
</table>

\(^a\)BDD: body dysmorphic disorder.
\(^b\)1=US $1.11.

**Figure 1.** Prevalence of body dysmorphic disorder among subsamples. BDD: body dysmorphic disorder.
Figure 2. Body parts of concern in our study population.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Gender</th>
<th>Univariate risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>284/1421 (20)</td>
<td>4.55 (3.14-6.58)</td>
<td>.001</td>
<td>3.02 (2.01-4.53)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>34/653 (5.2)</td>
<td>Ref&lt;sup&gt;a&lt;/sup&gt;</td>
<td>__b</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>4/17 (23.5)</td>
<td>5.60 (1.73-18.09)</td>
<td>.004</td>
<td>4.57 (1.28-16.38)</td>
</tr>
<tr>
<td>Mean age</td>
<td></td>
<td></td>
<td>0.93 (0.92-0.95)</td>
<td>.001</td>
<td>0.97 (0.95-0.99)</td>
</tr>
<tr>
<td>Age categories (years)</td>
<td>18-24</td>
<td>267/1091 (24.5)</td>
<td>2.69 (1.79-4.02)</td>
<td>.001</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>25-44</td>
<td>30/279 (10.8)</td>
<td>Ref&lt;sup&gt;a&lt;/sup&gt;</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>45-64</td>
<td>23/654 (3.5)</td>
<td>0.30 (0.17-0.53)</td>
<td>.001</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>&gt;64</td>
<td>2/66 (3)</td>
<td>0.26 (0.060-1.11)</td>
<td>.07</td>
<td>__</td>
</tr>
<tr>
<td>Occupation</td>
<td>Student</td>
<td>263/1043 (25.2)</td>
<td>Ref&lt;sup&gt;a&lt;/sup&gt;</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>Worker</td>
<td>50/865 (5.8)</td>
<td>0.18 (0.13-0.25)</td>
<td>.001</td>
<td>0.51 (0.28-0.94)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>9/183 (4.9)</td>
<td>0.15 (0.07-0.30)</td>
<td>.001</td>
<td>0.35 (0.14-0.88)</td>
</tr>
<tr>
<td>Dermatologic disease</td>
<td>Yes</td>
<td>204/1072 (19)</td>
<td>1.79 (1.40-2.29)</td>
<td>.001</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>118/1019 (11.6)</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Acne</td>
<td>Yes</td>
<td>115/532 (21.6)</td>
<td>1.80 (1.40-2.32)</td>
<td>.001</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>207/1559 (13.3)</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Atopic dermatitis</td>
<td>Yes</td>
<td>99/434 (22.8)</td>
<td>1.90 (1.46-2.48)</td>
<td>.001</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>223/1657 (13.5)</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Other dermatitis</td>
<td>Yes</td>
<td>10/35 (28.6)</td>
<td>2.23 (1.06-4.70)</td>
<td>.03</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>312/2056 (15.2)</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>Yes</td>
<td>2/50 (4)</td>
<td>0.22 (0.05-0.92)</td>
<td>.02</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>320/1041 (15.7)</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Psychiatric disorder</td>
<td>Yes</td>
<td>150/493 (30.4)</td>
<td>3.63 (2.83-4.65)</td>
<td>.001</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>172/1598 (10.8)</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Depressive Disorder</td>
<td>Yes</td>
<td>84/244 (34.4)</td>
<td>3.55 (2.64-4.78)</td>
<td>.001</td>
<td>1.68 (1.06-2.69)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>238/1847 (12.9)</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Eating behavior disorder</td>
<td>Yes</td>
<td>72/136 (52.9)</td>
<td>7.67 (5.34-11.02)</td>
<td>.001</td>
<td>3.77 (2.26-6.30)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>250/1955 (12.8)</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Obsessive-compulsive disorder</td>
<td>Yes</td>
<td>12/33 (36.4)</td>
<td>3.22 (1.57-6.62)</td>
<td>.001</td>
<td>__</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>310/2058 (15.1)</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
</tbody>
</table>
Given the heterogeneity of the population, most of whom were women aged 18-24 years, we performed a multivariate analysis considering gender and age (Table 3). On the one hand, after multivariate analysis by gender in the group of women, students, eating disorder, and anxiety disorder remained with a significance at <.05 as diagnostic predictors of BDD (Table 3). In the male group after multivariate analysis, none of the factors analyzed showed statistical significance. On the other hand, in the multivariate analysis by age (Table 3), the following factors remained with significance at <.05 as diagnostic predictors of BDD. In the 18-24 years analysis, the diagnostic predictors were female gender and other gender, students, depressive disorder, eating disorder, and anxiety disorder. In the 25-44 years analysis, the diagnostic predictor was income level between €2000 and €2999. In the 45-64 years analysis, the diagnostic predictor was female gender, and the >64 years analysis was not performed due to the small sample size.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Univariate risk ratio (95% CI)</th>
<th>P value</th>
<th>Multivariate risk ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attention-deficit/hyperactivity disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18/59 (30.5)</td>
<td>2.49 (1.41-4.40)</td>
<td>.003</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>304/2032 (15)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Anxiety disorder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>108/292 (37)</td>
<td>4.35 (3.30-5.74)</td>
<td>.001</td>
<td>2.30 (1.36-3.88)</td>
</tr>
<tr>
<td>No</td>
<td>214/1799 (11.9)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Plastic surgery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>57/355 (16.1)</td>
<td>1.06 (0.78-1.45)</td>
<td>.71</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>298/355 (83.9)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

**a**Ref: reference value.

**b**Not applicable.
<table>
<thead>
<tr>
<th>Variable</th>
<th>18-24 years age group</th>
<th></th>
<th></th>
<th>25-44 years age group</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BDDa prevalence, n/N (%)</td>
<td>Risk ratio (95% CI)</td>
<td>P value</td>
<td>Multivariate risk ratio (95% CI)</td>
<td>P value</td>
<td></td>
</tr>
<tr>
<td>Female gender</td>
<td>Occupation</td>
<td></td>
<td></td>
<td>Occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>occupation</td>
<td>Student</td>
<td>233/752 (30.9)</td>
<td>Refb</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>occupation</td>
<td>Worker</td>
<td>43/541 (7.9)</td>
<td>0.19 (0.14-0.27)</td>
<td>.001</td>
<td>0.46 (0.22-0.96)</td>
</tr>
<tr>
<td></td>
<td>occupation</td>
<td>Other</td>
<td>9/128 (7)</td>
<td>0.17 (0.08-0.34)</td>
<td>.001</td>
<td>0.26 (0.09-0.72)</td>
</tr>
<tr>
<td>Monthly income (€)d</td>
<td></td>
<td>&lt;500</td>
<td>181/617 (29.3)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500-999</td>
<td>30/107 (28)</td>
<td>0.94 (0.59-1.48)</td>
<td>.79</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000-1999</td>
<td>34/255 (13.3)</td>
<td>0.37 (0.25-0.55)</td>
<td>.001</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000-2999</td>
<td>8/192 (4.2)</td>
<td>0.10 (0.05-0.22)</td>
<td>.001</td>
<td>0.37 (0.15-0.93)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;3000</td>
<td>91/173 (5.2)</td>
<td>0.13 (0.07-0.26)</td>
<td>.001</td>
<td>—</td>
</tr>
<tr>
<td>Eating behavior disorder</td>
<td></td>
<td>Yes</td>
<td>69/125 (55.2)</td>
<td>6.195 (4.23-9.07)</td>
<td>.001</td>
<td>4.03 (2.32-6.98)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>215/1296 (16.6)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td></td>
<td>Yes</td>
<td>103/250 (41.2)</td>
<td>3.83 (2.85-5.16)</td>
<td>.001</td>
<td>2.94 (1.65-5.25)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>181/171 (15.5)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>18-24 years age group</td>
<td>Gender</td>
<td>Female</td>
<td>236/787 (30)</td>
<td>4.20 (2.75-6.42)</td>
<td>.001</td>
<td>3.19 (2.05-4.97)</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>Male</td>
<td>27/292 (9.2)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>Other</td>
<td>4/12 (33.3)</td>
<td>4.91 (1.39-17.37)</td>
<td>.01</td>
<td>6.38 (1.72-23.67)</td>
</tr>
<tr>
<td>Dermatologic disease</td>
<td>Occupation</td>
<td>Student</td>
<td>255/1005 (25.4)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Occupation</td>
<td>Worker</td>
<td>8/67 (11.9)</td>
<td>0.40 (0.19-0.85)</td>
<td>.02</td>
<td>0.39 (0.17-0.86)</td>
</tr>
<tr>
<td>Psychiatric disorder</td>
<td></td>
<td>Yes</td>
<td>177/640 (27.7)</td>
<td>1.53 (1.15-2.05)</td>
<td>.004</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>90/451 (20)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Depressive Disorder</td>
<td></td>
<td>Yes</td>
<td>130/301 (43.2)</td>
<td>3.62 (2.70-4.86)</td>
<td>.001</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>137/790 (17.3)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Eating disorder</td>
<td></td>
<td>Yes</td>
<td>73/152 (48)</td>
<td>3.55 (2.49-5.06)</td>
<td>.001</td>
<td>1.71 (1.04-2.84)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>194/930 (20.7)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Anxiety disorder</td>
<td></td>
<td>Yes</td>
<td>66/107 (61.7)</td>
<td>6.27 (4.12-9.54)</td>
<td>.001</td>
<td>3.70 (2.13-6.42)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>201/984 (20.4)</td>
<td>Ref</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: a BDD = Body Image Disturbance Disorder, b Ref = Reference category, c P value not provided.
Regarding quality of life, BDD was not statistically associated with physical health status. However, those diagnosed with BDD showed significantly lower levels of mental health than those without BDD (Figure 3). Additionally, quality of life analysis was performed considering psychiatric comorbidity as a factor affecting quality of life. Differences regarding mental health status remained statistically significant ($P < .05$) for patients with BDD, irrespective of the presence of psychiatric pathology (Figure 4).

### Figure 3. Association of body dysmorphic disorder with (A) physical and (B) mental health status. BDD: body dysmorphic disorder.
Discussion

Principal Findings

Our study reveals many interesting aspects of BDD. Most of our findings are consistent with those reported previously [1,4,6,12,13]. For example, BDD is particularly prevalent in the young adult population [5], and patients with BDD are particularly concerned about an average of 4.6 different body parts. However, some findings are novel in our study. For instance, the prevalence of BDD in our sample population (15.2%) was higher than expected (0.5%-3.2%) [1]. Further, in addition to the described association between BDD and eating and depressive disorders, we report that BDD is closely associated with anxiety disorders. The most relevant finding of our study is the perception of quality of life: participants with BDD had a poorer perception of quality of life associated with mental but not physical health problems. Moreover, the perception of quality of mental health in patients with BDD was independent of diagnosis of any mental disorder.

Our study reports that the prevalence of BDD in adults is 15.2% in Spain, which is higher than that reported in the general population (0.5%-3.2%) in another study [1] and higher than that reported in a Spanish multicentric study in patients with acne (10.6%) [2]. Our findings may be explained by 2 factors. First, our sample population was particularly young, with more than half of the participants being in the age group of 18-24 years. As shown in a previous study [6], the younger the age of the patient, the greater is the possibility of BDD diagnosis. Second, our data reflect more of a screening diagnosis as compared with a definitive diagnosis established with BDDQ via an interview by a health professional, which is more demanding [17].

Our study shows that the sociodemographic characteristics most associated with the diagnosis of BDD are gender (female and other), age group of 18-24 years, students, income level of less than €500/month, and a diagnosis of previous dermatologic or psychiatric disease. Regarding gender, females showed a statistically significant ($P<.001$) higher prevalence of BDD than males (284/1421, 19.9% vs 34/653, 5.2%, respectively). Previous studies have reported similar prevalence between men and women [3] or increased prevalence in women (68.5%) compared to that in men [4]. In our study, almost 67.9% (1421/2091) of the participants were women, which may suggest the need for future studies controlled by sex to clarify the differences. Regarding age, the mean age at diagnosis of BDD in the participants in our study was 23.5 years, which is higher than the mean age of 16.9 years described elsewhere [5]. This is probably because our study only included populations 18
years and older, resulting in an increase in the mean age at diagnosis. Furthermore, we found a higher prevalence of BDD among students (263/1043, 25.2%), which was also higher than 1.3%-5.8% reported in another study [1]. Again, this could be because almost half our sample was comprised of students. Moreover, this can be attributed to the influence of social media in the current age, increasing young people’s concern about their body image [1]. It would be appropriate to conduct a more specific study including this young population group.

Participants with BDD in our study were concerned about an average of 4.6 different body parts, which is in line with that reported by previous studies (5-7 body parts) [13]. The body parts that were of the most concern were body fat (248/322, 77%), thighs (191/322, 59.3%), face (166/322, 51.6%), hip (144/322, 44.7%), and skin (126/322, 39.1%). Skin (14/26, 54%) and face (8/26, 31%) were the body parts of the greatest concern in previous studies [1]. The prevalence of BDD in participants with dermatologic conditions (204/1072, 19%) in our study falls within the range (4.9%-21.1%) reported in the literature [1]. However, it is necessary to emphasize that previous studies did not distinguish between dermatologic patients per se and those undergoing aesthetic procedures [1].

Regarding the association between previous dermatologic disease and BDD, we found no significant relationships. The association between BDD and acne (115/532, 21.6%) in our study was slightly higher than that previously described (8.8%-21.1%) [1]. The association of BDD with having undergone rhinoplasty was 12.1% (39/322), which was lower than that previously documented (20.1%) [1]. However, the sample size limited our capability to extract meaningful conclusions on this issue. Among participants with previous psychiatric pathology and BDD, there was a significant association with eating disorders (72/136, 52.9%) and depressive disorders (84/244, 34.4%), similar to the findings of 12%-45% and 47%-56.3%, respectively, reported in a previous systematic review [1]. In addition, our study showed a significant comorbidity with anxiety disorders (108/292, 36.9%), making it necessary to conduct future studies in this subgroup.

Regarding quality of life, participants diagnosed with BDD had a perception of having a poorer mental health status than those without BDD. In contrast, there were no significant differences in the physical status between participants with and without BDD [12]. Ultimately, the diagnosis of BDD was associated with a perception of reduced quality of life that is not subsidiary to the presence of mental health disorders. In other words, our study suggests that BDD could be used as a marker or predictor of an individual’s perception of quality of life, which is independent of the presence of mental health problems.

Limitations and Strengths

Our study was based on the use of a questionnaire that was disseminated telematically, which is why we obtained a heterogeneous participant base, being represented mostly by female students in an age range of 18-24 years. The higher percentage of prevalence obtained may be linked to the specific population in this study. However, we must bear in mind that, despite obtaining the first diagnostic approximation with BDDQ, its confirmation must subsequently be backed up by an interview with a health professional [17]. Furthermore, the mean age at BDD diagnosis (23.5 years) in our study was higher than that (16.9 years) described previously [5] because this study was limited to participants aged 18 years and older. Additionally, the number of participants in our study who had undergone previous plastic surgeries was too low to achieve a proper statistical power.

With regard to the strengths of this study, we increased the number of variables related to BDD compared to the number of variables used in previous studies [2-6,12,13], which, together with the total number of participants, resulted in a large database. Further, we introduced the quality of life indicators through the SF-12v2 health survey.

Conclusion

Patients with BDD experience serious biopsychosocial repercussions. This study provides the first approximation of the prevalence of BDD in the Spanish population, which was found to be higher than expected, although our results should be interpreted cautiously. BDD was particularly prevalent in participants aged 18-24 years, students, and women. BDD is associated with psychiatric conditions such as eating disorder, anxiety disorder, and depressive disorder, and with dermatologic conditions such as acne and dermatitis. No significant associations were found between BDD and the performance of previous aesthetic procedures, which could be due to our small sample size. Finally, BDD could be a marker of an individual’s perception of quality of health, irrespective of psychiatric diagnoses. Future studies should confirm our preliminary findings.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors’ Contributions

HBF and BRA conceptualized this study, contributed to the methodology, supervised this study, and contributed to the resources. AL and CL curated the data and contributed to visualization. CL and MBF performed a formal analysis of this study. AL, HBF, and BRA contributed to the investigations. HBF contributed to project administration. AL and MBF wrote the original draft. HBF performed reviewing and editing.
Conflicts of Interest

In the last 24 months, HBF has received lecture fees from Takeda, BIAL, laboratorios Rubio, and laboratorios Rovi. He has also been granted 3 prizes for the development of a serious videogame for treating attention-deficit/hyperactivity disorder (The Secret Trail of Moon) called as the Shibuya Prize by Takeda, first prize of the College of Psychologists of Madrid, and a prize for the Best Innovative Health Initiative within Healthstart. He is the Principal Investigator of an IPFIS research contract (accessed on August 12, 2022; IF16/00039), Coprincipal Investigator of a MINECO research grant (RTI2018-101857-B-I00), and Principal Investigator of a research of the Sincronia project, funded by the start-up Bitsphi; recipient of a Fund for the Improvement of Postsecondary Education grant and an IDIPHIPA intensification grant; involved in 2 clinical trials (Mensia Koala, Newrofeed Study; ESKETSUI2002); and a cofounder of Haglaia Solutions. He is also an employee and member of the advisory board of Ita Salud Mental (Korian).

References


https://formative.jmir.org/2024/1/e46515 JMIR Form Res 2024 | vol. 8 | e46515 p.24 (page number not for citation purposes)


**Abbreviations**

- **BDD**: body dysmorphic disorder
- **BDDQ**: Body Dysmorphic Disorder Questionnaire
- **SF-12v2**: 12-item Short Form version 2
Methodological Insights on Recruitment and Retention From a Remote Randomized Controlled Trial Examining the Effectiveness of an Alcohol Reduction App: Descriptive Analysis Study

Melissa Oldham¹, PhD; Larisa Dinu¹, MSc; Gemma Loebenberg¹, MSc; Matt Field², PhD; Matthew Hickman³, PhD; Susan Michie⁴, PhD; Jamie Brown¹, PhD; Claire Garnett¹, PhD

¹University College London, London, London, United Kingdom
²Department of Psychology, University of Sheffield, Sheffield, United Kingdom
³Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
⁴Centre for Behaviour Change, University College London, London, United Kingdom

Corresponding Author:
Melissa Oldham, PhD
University College London
London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 20 7679 883
Email: m.oldham@ucl.ac.uk

Abstract

Background: Randomized controlled trials (RCTs) with no in-person contact (ie, remote) between researchers and participants offer savings in terms of cost and time but present unique challenges.

Objective: The goal of this study is to examine the differences between different forms of remote recruitment (eg, National Health Service [NHS] website, social media, and radio advertising) in the proportion of participants recruited, demographic diversity, follow-up rates, and cost. We also examine the cost per participant of sequential methods of follow-up (emails, phone calls, postal surveys, and postcards). Finally, our experience with broader issues around study advertising and participant deception is discussed.

Methods: We conducted a descriptive analysis of 5602 increasing-and-higher-risk drinkers (Alcohol Use Disorders Identification Test score ≥8), taking part in a 2-arm, parallel group, remote RCT with a 1:1 allocation, comparing the intervention (Drink Less app) with usual digital care (NHS alcohol advice web page). Participants were recruited between July 2020 and March 2022 and compensated with gift vouchers of up to £36 (a currency exchange rate of £1=US $1.26988 is applicable) for completing follow-up surveys, with 4 stages of follow-up: email reminders, phone calls, postal survey, and postcard.

Results: The three main recruitment methods were advertisements on (1) social media (2483/5602, 44.32%), (2) the NHS website (1961/5602, 35.01%), and (3) radio and newspapers (745/5602, 13.3%), with the remaining methods of recruitment accounting 7.37% (413/5602) of the sample. The overall recruitment cost per participant varied from £0 to £11.01. Costs were greater when recruiting participants who were men (£0–£28.85), from an ethnic minority group (£0–£303.81), and more disadvantaged (£0–£49.12). Targeted approaches were useful for recruiting more men but less useful in achieving diversity in ethnicity and socioeconomic status. Follow-up at 6 months was 79.58% (4458/5602). Of those who responded, 92.4% (4119/4458) responded by email. Each additional stage of follow-up resulted in an additional 2-3 percentage points of the overall sample being followed up, although phone calls, postal surveys, and postcards were more resource intensive than email reminders.

Conclusions: For remote RCTs, researchers could benefit from using a range of recruitment methods and cost-targeted approaches to achieve demographic diversity. Automated emails with substantial financial incentives for prompt completion can achieve good follow-up rates, and sequential, offline follow-up options, such as phone calls and postal surveys, can further increase follow-up rates but are comparatively expensive. We also make broader recommendations focused on striking the right balance when designing remote RCTs. Careful planning, ongoing maintenance, and dynamic decision-making are required throughout a trial to balance the competing demands of participation among those eligible, deceptive participation among those who are not eligible, and ensuring no postrandomization bias is introduced by data-checking protocols.
Introduction

Randomized controlled trials (RCTs) are used to examine the efficacy of interventions on a wide range of health-related behaviors and outcomes [1-4]. RCTs examining the efficacy of digital interventions are increasingly taking place on the web or remotely. Web-based trials feature no in-person contact between researchers and participants, with the administration of the intervention and all measures completed on the web. Remote trials also have no in-person contact between researcher and participant but may involve some offline follow-up options, such as completing surveys over the phone or by post. Web-based and remote trials can be cheaper and less labor-intensive than in-person trials, although they present some unique challenges around recruitment, retention, and participant deception. Here, we present methodological insights from a large-scale (n=5602) remote RCT examining the effectiveness of a digital intervention, the “Drink Less” app [5], in helping increasing-and-higher-risk drinkers (Alcohol Use Disorders Identification Test [AUDIT] score ≥8) to reduce their alcohol consumption.

Digital interventions, such as websites and apps, are increasingly being used for a wide range of health behaviors [6] and can offer benefits over in-person interventions in terms of cost, convenience, and anonymity [7]. RCTs aiming to evaluate digital interventions can be conducted on the web or remotely and may have several advantages relative to trials requiring in-person contact. First, web-based and remote settings could increase the external validity of the trial, as having to travel to in-person appointments for baseline or follow-up assessments does not reflect real-world implementation or how users access digital interventions [8]. Second, in theory, participants can be recruited from throughout nations or even globally, giving a larger and potentially more generalizable sampling frame [9]. Third, the cost of web-based or remote trials is likely to be much less as they can be partly automated, reducing demands on researcher time, and could potentially reduce researcher bias through double blinding [9].

However, there are also significant challenges with web-based or remote RCTs beyond those conducted in person. First, it may be harder to recruit participants or to recruit a broadly representative sample [10,11], as some groups, such as older adults and people from less advantaged communities, may be less likely to engage with research conducted remotely [12]. Second, researchers have less control over who signs up, and it is possible that motivated individuals may sign up multiple times for financial incentives [8,13]. Third, once recruited, researchers may have less control over how participants engage with the intervention [14] or respond to follow-ups [8]. This could be particularly problematic with groups who may have low digital literacy and may not understand how to use the intervention, although this may be reflective of how people would engage with digital interventions in real-world settings. There are other challenges that are present in both remote and in-person trials. Contamination occurs when the comparator group finds the intervention being tested outside of the trial [9]. This could be particularly likely if the comparator group receives an intervention they do not deem acceptable and seeks out alternatives. These biases could introduce bias into RCTs, which could obscure the effect of the intervention.

Here, we draw on data from a large-scale remote RCT, evaluating the effectiveness of the Drink Less app [5] compared with usual digital care (the National Health Service [NHS] alcohol advice web page). Drink Less is a theory- and evidence-informed, app-based intervention designed by researchers [15,16] to help increasing-and-higher-risk drinkers reduce their alcohol consumption. To mitigate some of the potential challenges outlined above, the trial used a multipronged recruitment strategy, including an advertisement on the NHS website and social media advertising [5]. In line with previous research [11], and to maximize follow-up rates, we offered substantial financial incentives to complete follow-up surveys, including an additional amount for completing the primary outcome within the first 24 hours, and undertook a comprehensive follow-up approach by sequentially sending follow-up reminders through email, SMS text messages, and telephone and by post. These strategies and broader methodological issues will be discussed ahead.

This study aims to:

1. Compare different remote recruitment methods in terms of cost per recruited participant, retention rates, participant deception, and sociodemographic diversity.
2. Compare the proportion of returned responses using different strategies for follow-up at each time point, and compare the cost and time associated with each follow-up stage.
3. Consider broader methodological issues pertaining to recruitment, retention, and participant deception, and discuss the success of strategies to mitigate these issues throughout the trial.

Methods

The protocol and analysis plan were preregistered on the Open Science Framework [17]. The trial was registered (ISRCTN64052601). The main trial findings are reported elsewhere [18].

Design

Participants
A total of 5602 participants were randomized in the RCT evaluating Drink Less. Participants were eligible if they were aged 18 years or older, lived in the United Kingdom, were increasing-and-higher-risk drinkers (AUDIT score ≥ 8), had access to an iOS device (iPhone, iPod touch, or iPad), and wanted to drink less alcohol. Recruitment ran from July 2020 to March 2022 and included an advertisement on the NHS website, a mail-out to a database of UK-based users of the smoking cessation app “Smoke Free”, radio and social media advertising, press releases, and local advertising through health care providers. Advertisements were codeveloped with public representatives.

Informed consent was sought at baseline to participate in 3 web-based follow-up surveys at 1, 3, and 6 months. Surveys were completed on the web through Qualtrics (Silver Lake), although at the 6-month follow-up, offline options (eg, phone and post) were available. The 6-month follow-up survey assessed primary and secondary outcomes relating to alcohol use and a range of related measures. The 1- and 3-month follow-up surveys only assessed secondary outcome measures relating to alcohol use. We attempted to contact participants within 30 days of their first invitation to complete each follow-up survey. To maximize data retention and to allow for time taken for answers to be posted at 6-month follow-up, data provided up to 2 weeks after the 30-day period were accepted.

Initially, as well as through 3 emails (days 0, 5, and 9) and (from January 15, 2022) a total of 2 SMS text messages (days 5 and 9), we had planned that at the 1-, 3-, and 6-month follow-up, all participants would also be sequentially offered opportunities to complete follow-up through phone (called twice from days 10 to 17), a mailed survey (from day 18), and a mailed postcard (from day 30). However, due to resource constraints, from November 2020 on, we only used automated emails on days 0, 5, 9, and 11 to contact participants at the 1- and 3-month follow-up; we no longer called or sent postal surveys. With the aim of improving these follow-up rates with less resource, we added SMS text messaging follow-ups. Phone calls, mailed surveys, and postcard follow-ups were retained for the 6-month follow-up survey (when the primary outcome was measured).

Measures
Recruitment Method
At baseline, participants were asked to specify where they saw the study advertised, with the following response options: NHS website, social media (eg, Facebook and Twitter [subsequently rebranded X]), other media (eg, radio and newspapers), emailed by the Smoke Free app, local health care provider, word of mouth, Google, general practitioner (GP) surgery, or other. If they selected “other,” free-text responses that fell within one of the response options were recoded (eg, Facebook would be social media). The response options “local health care provider” and “GP surgery” were collapsed. Throughout the study, both untargeted and targeted (eg, at men) social media advertisements were used. These were analyzed separately.

Participant Deception
We experienced 3 distinct subgroups of participant deception throughout the trial: duplicates, manual fraud, and bots. Duplicate responses, where individuals signed up more than once with identical names and phone numbers, were the least prevalent (n=49) and easiest to detect. Data checks were undertaken each month to search for duplicate values. Manual fraud was a more prevalent form of participant deception (n=297), defined as individuals who signed up multiple times with false information, such as phone numbers linked to businesses where they were not known or addresses that did not exist. To identify manual fraud, monthly checks were made on all addresses and telephone numbers provided to ensure street names matched the postcode and that numbers were mobile phone numbers. Any suspicious responses were flagged, and the participants were contacted and asked to confirm their details over the phone. Where individuals were not known at the phone number provided, they were removed from the study. To make it easier to automatically screen out those engaging in manual fraud, we added attention checks, whereby individuals were asked to select a certain response option. Participants were also asked to confirm their age at 2 different points in the baseline survey to ensure they were consistent. Individuals failing either of these attention checks were screened out of the survey before randomization. The most prevalent type of fraud were “bot” responses (n=863). These were fraudulent responses similar to manual fraud, but they occurred in batches of 20-30 at a time when contact information was given in noticeably similar formats (eg, firstname123@emailaddress.com), often with American street addresses (being UK-based was an inclusion criteria of the trial). These responses seemed to be automated and were identified using the same process of address checking as above (individuals not known at the phone number provided were removed from the study). Adding a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) to the survey eliminated this issue. A more detailed discussion on participant deception is described elsewhere [20].

Sociodemographic Characteristics
Sociodemographic measures were assessed at baseline. This study focuses on gender, ethnicity, and occupation (to derive socioeconomic status [SES]: ABC1 [managerial, professional, and intermediate occupations] versus C2DE [skilled, semiskilled, unskilled manual, and lowest-grade worked or unemployed]).

Analysis
Aim 1: Methods of Recruitment
Each recruitment method is compared in terms of the proportion of enrolled participants, the proportion of participants who were men, from a minority ethnic group, or from a more disadvantaged background (C2DE), and the proportion of participant deception. Cost-per-recruited participant citing each recruitment method (eg, total spend on recruitment method divided by the number of participants citing recruitment method) is reported. As well as the overall cost per participant, we also present the cost per participant stratified by gender (eg, for each
man recruited), ethnicity (eg, for those from ethnic minority individuals), and SES (eg, for those from more disadvantaged backgrounds). Finally, we present follow-up rates at 1-, 3-, and 6-months for each method of recruitment.

**Aim 2: Follow-Up**

The proportion of the sample responding at each sequential stage of follow-up (ie, emails, phone calls, postal surveys, and postcards) is reported. The cost of each follow-up stage per participant responding at each stage is also reported. This was derived by dividing the estimated researcher time and other relevant costs by the number of follow-ups completed at each stage.

**Aim 3: Broader Methodological Issues**

Broader methodological issues such as advertising, participant deception, technical support, contamination, and boosting retention are discussed. We describe and briefly discuss the strategies we used throughout the trial to mitigate issues.

**Ethical Considerations**

Ethical approval for the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone app) trial was granted by the University College London (UCL) Ethics Committee (16799/001). Participants provided informed consent before participating in the trial. Study data were pseudonymized and stored on a secure university drive. Participants were compensated with gift vouchers of up to £36 (a currency exchange rate of £1=US $1.26988 is applicable) for completing the 3 surveys: £6 for the survey at 1 and 3 months and £12 at 6 months, with an additional £12 if the 6-month survey was completed within 24 hours.

**Results**

**Sample Characteristics**

A total of 5602 participants completed the baseline survey between July 2020 and March 2022: 65.78% (3685/5602) responded at 1-month follow-up, 63.80% (3574/5602) at 3-month follow-up, and 79.58% (4458/5602) at 6-month follow-up. Over half (3207/5602, 57.25%) of the sample were women, 42.22% (2365/5602) were men, 0.46% (26/5602) were “other,” and 0.07% (4/5602) preferred not to say. Most of the sample were White (5296/5602, 94.54%) and earned above-average income (4151/5602, 74.01%). The sample characteristics were similar at each follow-up. Table 1 reports the sociodemographic characteristics of the sample at baseline and those responding at each stage of follow-up.

**Table 1. Sample characteristics at baseline and among those who responded at 1-month, 3-month, and 6-month follow-up for increasing-and-higher-risk drinkers participating in the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial (RCT).**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Baseline a (N=5602)</th>
<th>1-month follow-up (n=3685)</th>
<th>3-month follow-up (n=3574)</th>
<th>6-month follow-up b (n=4458)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>3207 (57.25)</td>
<td>2046 (55.52)</td>
<td>1992 (55.74)</td>
<td>2534 (56.84)</td>
</tr>
<tr>
<td>Women</td>
<td>2365 (42.22)</td>
<td>1620 (43.96)</td>
<td>1565 (43.79)</td>
<td>1903 (42.69)</td>
</tr>
<tr>
<td>Other</td>
<td>26 (0.46)</td>
<td>16 (0.43)</td>
<td>14 (0.39)</td>
<td>17 (0.38)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>4 (0.07)</td>
<td>3 (0.08)</td>
<td>3 (0.08)</td>
<td>4 (0.09)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>96 (1.71)</td>
<td>68 (1.85)</td>
<td>68 (1.9)</td>
<td>83 (1.86)</td>
</tr>
<tr>
<td>Black</td>
<td>47 (0.84)</td>
<td>35 (0.95)</td>
<td>39 (1.09)</td>
<td>41 (0.92)</td>
</tr>
<tr>
<td>Chinese</td>
<td>9 (0.16)</td>
<td>9 (0.24)</td>
<td>9 (0.25)</td>
<td>9 (0.2)</td>
</tr>
<tr>
<td>White</td>
<td>5296 (94.54)</td>
<td>3474 (94.27)</td>
<td>3361 (94.04)</td>
<td>4206 (94.35)</td>
</tr>
<tr>
<td>Mixed</td>
<td>113 (2.02)</td>
<td>75 (2.03)</td>
<td>71 (1.99)</td>
<td>84 (1.88)</td>
</tr>
<tr>
<td>Other</td>
<td>21 (0.37)</td>
<td>15 (0.41)</td>
<td>15 (0.42)</td>
<td>18 (0.4)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>19 (0.34)</td>
<td>9 (0.24)</td>
<td>11 (0.31)</td>
<td>16 (0.36)</td>
</tr>
<tr>
<td>Not known</td>
<td>1 (0.02)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (0.02)</td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC1 b</td>
<td>4151 (74.01)</td>
<td>2759 (74.87)</td>
<td>2688 (75.21)</td>
<td>3337 (74.85)</td>
</tr>
<tr>
<td>C2DE c</td>
<td>1451 (25.9)</td>
<td>926 (25.13)</td>
<td>886 (24.79)</td>
<td>1121 (25.15)</td>
</tr>
</tbody>
</table>

aThe data is also reported in the main trial paper [18].
bABC1: managerial, professional, and intermediate occupations.
cC2DE: skilled, semiskilled, unskilled manual, and lowest-grade worked or unemployed.
Aim 1: Recruitment Methods, Demographic Diversity, and Cost Per Participant

Most participants recruited for this trial reported seeing it advertised on social media (2483/5602, 44.32%), the NHS website (1961/5602, 35.01%), or through radio or newspapers (745/5602, 13.3%), with all other recruitment methods accounting for 7.37% (413/5602) of the sample (Table 2).

Table 2. Total recruitment and proportion of recruited sample of iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial who were men, of minority ethnic groups, had lower socioeconomic status (SES), and identified as a fraudulent response by recruitment method.

<table>
<thead>
<tr>
<th>Recruitment method</th>
<th>Included sample (N=5602), n (%)</th>
<th>Men, n (%)</th>
<th>Ethnic minority group, n (%)</th>
<th>Low SES, n (%)</th>
<th>Fraudulent response, n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untargeted social media</td>
<td>2119 (37.82)</td>
<td>650 (30.67)</td>
<td>147 (6.94)</td>
<td>507 (23.93)</td>
<td>1020/3139 (32.49)</td>
</tr>
<tr>
<td>Targeted social media</td>
<td>364 (6.5)</td>
<td>353 (96.98)</td>
<td>13 (3.57)</td>
<td>90 (24.73)</td>
<td>8/372 (2.15)</td>
</tr>
<tr>
<td>National Health Service website</td>
<td>1961 (35.01)</td>
<td>628 (32.02)</td>
<td>76 (3.88)</td>
<td>570 (29.07)</td>
<td>123/2084 (5.9)</td>
</tr>
<tr>
<td>Radio or newspapers</td>
<td>745 (13.3)</td>
<td>591 (79.33)</td>
<td>27 (3.62)</td>
<td>167 (22.42)</td>
<td>19/764 (2.49)</td>
</tr>
<tr>
<td>Word of mouth</td>
<td>142 (2.53)</td>
<td>74 (52.11)</td>
<td>9 (6.34)</td>
<td>41 (28.87)</td>
<td>11/153 (7.19)</td>
</tr>
<tr>
<td>Google</td>
<td>159 (2.84)</td>
<td>44 (27.67)</td>
<td>7 (4.4)</td>
<td>50 (31.45)</td>
<td>9/168 (5.56)</td>
</tr>
<tr>
<td>Smoke Free email</td>
<td>55 (0.98)</td>
<td>13 (23.64)</td>
<td>3 (5.45)</td>
<td>11 (20)</td>
<td>10/65 (15.38)</td>
</tr>
<tr>
<td>Health care provider or general practitioner</td>
<td>15 (0.27)</td>
<td>5 (33.33)</td>
<td>3 (20)</td>
<td>4 (26.67)</td>
<td>16/31 (51.61)</td>
</tr>
<tr>
<td>Other</td>
<td>42 (0.75)</td>
<td>7 (16.67)</td>
<td>2 (4.76)</td>
<td>11 (26.19)</td>
<td>0/42 (0)</td>
</tr>
</tbody>
</table>

aThe percentage of participants recruited from each method (ie, N=Included sample value).
bThe percentage of the overall sample from each recruitment method including those removed after participant deception checks.

Ongoing sociodemographic tracking throughout the study revealed that women, White, and advantaged participants were being overrecruited. In response, strategies targeted at a more diverse sample in terms of gender, SES, and ethnicity were introduced with mixed success. This included targeted social media advertisements aimed at men and radio advertisements on Talk Radio, Asian Sounds (in English and Urdu), and Punjabi Radio (in English and Punjabi).

Recruitment methods differed in the proportion of men (range 17%-97%), with targeted approaches including social media advertising (353/364, 97% men) and radio advertising (591/745, 79.3%) being the most successful in recruiting a sample of men. Word of mouth was most effective in terms of recruiting a balanced sample in terms of gender (74/142, 52.1% men) but recruited a small proportion (142/5602, 2.53%) of the sample overall.

Recruiting through GP surgeries and local health care providers resulted in the highest proportion of participants from minority ethnic groups (3/15, 20%) but recruited a small proportion of participants in total (15/5602, 0.27%). Untargeted social media advertisements and word of mouth were the next best, with 6.94% (147/2119) and 6.3% (9/142) of the sample coming from ethnic minority individuals, respectively.

The NHS website, word of mouth, and Google all recruited around a third of participants who were more disadvantaged. However, both Google (159/5602, 2.84%) and word of mouth (142/5602, 2.53%) recruited a small proportion of participants in total.

The final column of Table 2 refers to the proportion (and number) of participants who were removed from the study due to participant deception, citing each recruitment method. A total of 84.54% (1028/1216) of participants identified as fraudulent cited social media as the place they saw the advertisement. It should be noted here that these participants may not have been honest in terms of where they saw the study advertisement and may have been deliberately misreporting where they found the study or responding at random.

Money spent on each of the recruitment methods varied from £0 for the NHS advertisement and word of mouth to £8203 for radio or newspaper advertisements (Table 3; a currency exchange rate of £1=US $1.26988 is applicable). Of the paid forms of recruitment, social media advertising and advertising through health care providers were the cheapest ways of recruiting participants who were men, of ethnic minorities, or from more disadvantaged backgrounds.

Although the overall number of participants recruited from health care settings was low, this was impeded by the COVID-19 pandemic. The initial recruitment plan was to have posters in primary care surgeries throughout the United Kingdom; however, due to the pandemic and associated lockdowns for most of the recruitment period, many people received health care on the web and were not visiting GP surgeries. We only started advertising in GP surgeries for the last 5 months of trial recruitment (in November 2021).

Those recruited from health care providers (15/15, 100%), Smoke Free email (51/55, 93%), and word of mouth (126/142, 88.7%) appeared to have the highest response rates and those recruited through advertisements on Google (109/159, 69%), and the NHS website (1513/1961, 77%) appeared among the lowest. Table 4 presents the follow-up rates at 1-, 3-, and 6-month follow-up.
Table 3. Total cost per participant and cost per participant who were men, of ethnic minority groups, and lower socioeconomic status (SES) by recruitment method for participants of the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial. A currency exchange rate of £1=US $1.26988 is applicable.

<table>
<thead>
<tr>
<th>Recruitment method</th>
<th>Recruited, n</th>
<th>Total cost (£)</th>
<th>Cost per participant (£)</th>
<th>Cost per man (£)</th>
<th>Cost per ethnic minority participant (£)</th>
<th>Cost per low SES participant (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untargeted social media</td>
<td>2119</td>
<td>6750.00</td>
<td>3.19</td>
<td>10.38</td>
<td>45.92</td>
<td>13.31</td>
</tr>
<tr>
<td>Targeted social media</td>
<td>364</td>
<td>690.00</td>
<td>1.90</td>
<td>1.95</td>
<td>53.08</td>
<td>7.67</td>
</tr>
<tr>
<td>National Health Service website</td>
<td>1961</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Radio or newspapers</td>
<td>745</td>
<td>8203.00</td>
<td>11.01</td>
<td>13.88</td>
<td>303.81</td>
<td>49.12</td>
</tr>
<tr>
<td>Word of mouth</td>
<td>142</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Google</td>
<td>159</td>
<td>1247.00</td>
<td>7.84</td>
<td>28.34</td>
<td>178.14</td>
<td>138.56</td>
</tr>
<tr>
<td>Smoke Free email</td>
<td>55</td>
<td>375.00</td>
<td>6.82</td>
<td>28.85</td>
<td>125.00</td>
<td>34.09</td>
</tr>
<tr>
<td>Health care provider or general practitioner</td>
<td>15</td>
<td>61.00</td>
<td>0.40</td>
<td>1.20</td>
<td>20.33</td>
<td>15.24</td>
</tr>
<tr>
<td>Other</td>
<td>42</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4. Follow-up rates at 1-, 3-, and 6-month follow-up among increasing-and-higher-risk drinkers participating in the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial by recruitment method.

<table>
<thead>
<tr>
<th>Recruitment method</th>
<th>Recruited, n</th>
<th>Follow-up rate at 1 month, n (%)</th>
<th>Follow-up rate at 3 months, n (%)</th>
<th>Follow-up rate at 6 months, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untargeted social media</td>
<td>2119</td>
<td>1376 (64.94)</td>
<td>1340 (63.24)</td>
<td>1708 (80.6)</td>
</tr>
<tr>
<td>Targeted social media</td>
<td>364</td>
<td>277 (76.1)</td>
<td>256 (70.33)</td>
<td>295 (81.04)</td>
</tr>
<tr>
<td>National Health Service website</td>
<td>1961</td>
<td>1237 (63.08)</td>
<td>1210 (61.7)</td>
<td>1513 (77.15)</td>
</tr>
<tr>
<td>Radio or newspapers</td>
<td>745</td>
<td>520 (69.8)</td>
<td>495 (66.44)</td>
<td>603 (80.94)</td>
</tr>
<tr>
<td>Word of mouth</td>
<td>142</td>
<td>107 (75.35)</td>
<td>106 (74.65)</td>
<td>126 (88.73)</td>
</tr>
<tr>
<td>Google</td>
<td>159</td>
<td>87 (54.72)</td>
<td>81 (50.94)</td>
<td>109 (68.55)</td>
</tr>
<tr>
<td>Smoke Free email</td>
<td>55</td>
<td>38 (69.09)</td>
<td>43 (78.18)</td>
<td>51 (92.73)</td>
</tr>
<tr>
<td>Health care provider or general practitioner</td>
<td>15</td>
<td>12 (80)</td>
<td>10 (66.66)</td>
<td>15 (100)</td>
</tr>
<tr>
<td>Other</td>
<td>42</td>
<td>31 (73.81)</td>
<td>33 (78.57)</td>
<td>38 (90.48)</td>
</tr>
</tbody>
</table>

Aim 2: Retention During Sequential Follow-Up

At 6-month follow-up, 92.4% (4119/4458) of those who responded did so in response to 1 of the 3 email notifications. An additional 2.02% (90/4458) responded following 2 phone calls from the research team, and 3.25% (145/4458) responded following a postal survey. The final stage of recruitment, a postcard sent through mail to participants featuring just the key outcome measure for the trial (AUDIT-C), yielded a further 2.33% (104/4458) of the followed-up sample. The estimated costs of each sequential stage of follow-up are presented in Tables 5 and 6.
Table 5. A summary of time spent and cost on different stages of contacting participants of the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trials at 1- and 3-month follow-up. A currency exchange rate of £1=US $1.26988 is applicable.

<table>
<thead>
<tr>
<th>Time point and method</th>
<th>Total follow-up sent, n</th>
<th>Follow-up per month, mean (range)</th>
<th>Responded</th>
<th>Hours spent sending follow-up(^a)</th>
<th>Hours spent sending vouchers</th>
<th>Total research hours</th>
<th>Cost research hours(^b) (£)</th>
<th>Other costs (£)</th>
<th>Total cost (£)</th>
<th>Cost per participant (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1-month follow-up</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated email</td>
<td>5602</td>
<td>267 (65-621)</td>
<td>1874</td>
<td>0</td>
<td>34</td>
<td>34</td>
<td>672</td>
<td>0</td>
<td>672</td>
<td>0.36</td>
</tr>
<tr>
<td>First manual email and SMS text message(^c)</td>
<td>3800 and 1057</td>
<td>181 (34-448)</td>
<td>1130</td>
<td>130</td>
<td>20</td>
<td>150</td>
<td>2966</td>
<td>106(^d)</td>
<td>3072</td>
<td>2.72</td>
</tr>
<tr>
<td>Second and third manual email and SMS text message(^c)</td>
<td>3291 and 648</td>
<td>175 (0-462)</td>
<td>643</td>
<td>112</td>
<td>12</td>
<td>124</td>
<td>2452</td>
<td>65(^d)</td>
<td>2517</td>
<td>3.91</td>
</tr>
<tr>
<td><strong>3-month follow-up</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated email</td>
<td>5602</td>
<td>267 (65-621)</td>
<td>2053</td>
<td>0</td>
<td>37</td>
<td>37</td>
<td>732</td>
<td>0</td>
<td>732</td>
<td>0.36</td>
</tr>
<tr>
<td>First manual email and SMS text message(^c)</td>
<td>3610 and 1282</td>
<td>172 (26-419)</td>
<td>1056</td>
<td>123</td>
<td>19</td>
<td>142</td>
<td>2807</td>
<td>128(^d)</td>
<td>2935</td>
<td>2.78</td>
</tr>
<tr>
<td>Second and third manual email and SMS text message(^c)</td>
<td>3698 and 874</td>
<td>176 (0-511)</td>
<td>460</td>
<td>126</td>
<td>8</td>
<td>134</td>
<td>2649</td>
<td>87(^d)</td>
<td>2736</td>
<td>5.95</td>
</tr>
</tbody>
</table>

\(^a\) Time spent sending manual reminders and SMS text messages. On average, an email and SMS text message reminder took 2 minutes and 5 seconds to send, and a voucher email took 1 minute and 8 seconds to send.

\(^b\) The cost here is the average of 2 research staff salaries (£19.77) \(\times\) research hours.

\(^c\) For the first 3 months of follow-up, we contacted participants twice manually by email, followed sequentially by phone calls, a written survey, and a postcard with the primary outcomes. However, this was not sustainable, so the subsequent follow-up stages were dropped at 1 and 3 months and are not presented below but are included in this total. 1-month phone completions=22, and 1-month postcard completions=16. 3-month phone completions=4, and 3-month postcard completions=1. SMS text messages were added 18 months into recruitment and sent at the same time as the first and second emails, so individual effects cannot be differentiated. SMS text messages did not add significantly to the time spent sending them, as they were also sent through mail merge at the same time.

\(^d\) Based on 10 pence (US $0.12) per SMS text message.

"
Table 6. A summary of time spent and cost on different stages of contacting participants of the iDEAS (iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone) randomized controlled trial (RCT) at 6-month follow-up. A currency exchange rate of £1=US $1.26988 is applicable.

<table>
<thead>
<tr>
<th>Method</th>
<th>Total follow-up sent, n</th>
<th>Follow-up per month, mean (range)</th>
<th>Responded (n=4458)</th>
<th>Follow-Up hours&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Voucher hours&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Data entry hours&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Total hours</th>
<th>Cost hours&lt;sup&gt;d&lt;/sup&gt; (£)</th>
<th>Other costs (£)</th>
<th>Total cost (£)</th>
<th>Cost per participant (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email through Qualtrics</td>
<td>5602</td>
<td>266 (64-621)</td>
<td>2358</td>
<td>0</td>
<td>42</td>
<td>e</td>
<td>42</td>
<td>830</td>
<td>0</td>
<td>830</td>
<td>0.35</td>
</tr>
<tr>
<td>First manual follow-up email and SMS text message&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1886 and 505</td>
<td>52 (4-132)</td>
<td>948</td>
<td>64</td>
<td>17</td>
<td>—</td>
<td>81</td>
<td>1601</td>
<td>51&lt;sup&gt;g&lt;/sup&gt;</td>
<td>1652</td>
<td>1.74</td>
</tr>
<tr>
<td>Second manual follow-up email and SMS text message&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1077 and 450</td>
<td>51 (4-132)</td>
<td>813</td>
<td>37</td>
<td>15</td>
<td>—</td>
<td>52</td>
<td>1028</td>
<td>45</td>
<td>1073</td>
<td>1.32</td>
</tr>
<tr>
<td>Phone calls</td>
<td>2118</td>
<td>101 (8-260)</td>
<td>90</td>
<td>117</td>
<td>2</td>
<td>—</td>
<td>119</td>
<td>2353</td>
<td>0</td>
<td>2353</td>
<td>26.14</td>
</tr>
<tr>
<td>Posted survey</td>
<td>1378</td>
<td>66 (2-167)</td>
<td>145</td>
<td>68</td>
<td>3</td>
<td>24</td>
<td>95</td>
<td>1878</td>
<td>2384&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4262</td>
<td>29.39</td>
</tr>
<tr>
<td>Postcard</td>
<td>1161</td>
<td>55 (2-156)</td>
<td>104</td>
<td>59</td>
<td>2</td>
<td>9</td>
<td>70</td>
<td>1384</td>
<td>1080&lt;sup&gt;i&lt;/sup&gt;</td>
<td>2464</td>
<td>23.69</td>
</tr>
</tbody>
</table>

<sup>a</sup>Based on average times of 2.05 minutes per email or SMS text message, 3.31 minutes per phone call, 2.94 minutes per survey, and 3.07 minutes per postcard.

<sup>b</sup>A voucher email took 1.08 seconds to send.

<sup>c</sup>Based on 10 minutes to input a survey and 5 minutes to input a postcard.

<sup>d</sup>The cost here is the average of 2 research staff salaries (£19.77) × research hours.

<sup>e</sup>Not available.

<sup>f</sup>Text messages were added 18 months (from January 15, 2022) into recruitment and sent at the same time as the first and second emails, so individual effects cannot be differentiated.

<sup>g</sup>Based on 10 pence (US $0.12) per SMS text message.

<sup>h</sup>Based on estimated stationary and postage costs of £1.73 per survey.

<sup>i</sup>Based on estimated stationary and postage costs of £0.93 per postcard.

### Aim 3: Broader Methodological Insights

#### Retention

Each SMS text message cost 10 pence (US $0.12) to send and required minimal researcher time as texts were sent to participants through mail merge at the same time as email reminders were sent. This was relatively low cost and low effort, and there was an increase in the average follow-up rate at 1- and 3-month follow-up in the 3 months before and after the introduction of the SMS text messages (from 58.0% (221/381) to 71.43% (830/1162) at 1 month and 58.5% (223/381) to 64.80% (753/1162) at 3 months).

#### Recruitment

Remote trials may unintentionally exclude participants with less experience using web-based surveys and digital interventions or with lower digital literacy. To mitigate this risk, in the recommendation email and at the end of the baseline survey, we included a link to a pictorial step-by-step guide to downloading and using the app [21] and encouraged participants to contact the research team if they needed technical support. Less than 10 participants contacted the research team for technical support throughout the trial.

### Advertisement Development

Advertising any research study involves balancing incentivizing the target audience to participate while avoiding incentivizing those outside of the target market to falsify information to gain reimbursement. This is particularly true of remote research, where there is no face-to-face contact with researchers and therefore fewer barriers to participant deception. Below, we outline the process of developing the study advertisement, involving feedback from public and patient involvement (PPI) groups and dynamic changes throughout the trial in response to higher rates of participant deception.

#### PPI Feedback on Advertising

To improve the clarity and appeal of the advertisement, we attended meetings with 2 PPI groups (the Sheffield Addiction Recovery Research Panel and the Alcohol and Food Discussion Group at the University of Stirling) and asked for feedback on an advertisement we had designed (Figure 1). The PPI group highlighted language (eg, “Researchers at UCL” and “trial”) that they felt was too formal and would make the study sound frightening or labor-intensive. Furthermore, they did not like the phrase “digital support tools,” which they felt was unclear, and instead suggested we use the phrase “online support tools.” The group also suggested that to make the advertisement more appealing, we should make it clear that people would get support...
to drink less alcohol, highlight the financial incentives in a more prominent position, and include pictures.

**Figure 1.** The original study advertisement designed by the research team to recruit participants to the iDEAS randomized controlled trial (left), advertisement following public and patient involvement feedback (middle), and advertisement following issues with participant deception (right). iDEAS: iOS Drink Less, Evaluating the Effectiveness of an Alcohol Smartphone; UCL: University College London.

### Advertising and Participant Deception

Following issues with participant deception, edits were made to the advertisement to disincentivize those who did not meet the inclusion criteria from signing up for financial reimbursement. The mention of the vouchers was removed from the heading and moved to the body of the advertisement. The specific amount was removed, and the text was updated to make it clear that there was no immediate financial incentive to participate in the study; rather, vouchers were sent after 1-, 3-, and 6-month follow-up surveys were completed.

### Negative Engagement With Advertising

Throughout the study, we also experienced negative engagement with our social media advertising, particularly on Facebook. Unhelpful comments included those joking about wanting to drink more (eg, “I need support with drinking MORE alcohol”), leaving negative messages about the research team (eg, “killjoy weirdos”), highlighting the reimbursement amount (eg, “vouchers sound good”), or telling other users reasons they had been screened out (eg, “people who use android rather than apple ones are not wanted”). We decided against disabling comments on advertising posts, as other people used them to engage positively with the study and to tag friends. Rather than respond to or delete posts, which may have further antagonized people, we used the “hide” feature on negative comments on a weekly basis, meaning these comments could not be seen by others but that the original poster was not notified. A total of 46.6% (210/451) of comments were hidden throughout the study.

### Contamination

This was a pragmatic trial, as we were testing the effect of the recommendation rather than the use of the Drink Less app. Nevertheless, we took steps to minimize contamination. We were careful not to mention the name of the app or the trial in any advertising. We also included 2 sensitivity analyses to try and capture the extent of contamination in the trial. One focused on those who followed the recommendation determined by self-report (at 1-month and 6-month follow-ups). The second was an instrumental variable analysis that accounted for nonuse in the intervention group and contamination in the comparator group by operationalizing the difference in app use between the 2 groups.

These recommendations are summarized in Textbox 1.
Textbox 1. Methodological recommendations for remote randomized controlled trials (RCTs).

Recruitment

- Use a range of recruitment methods.
- Monitor the demographic composition of the sample during trials and have targeted methods for underrecruited groups.
- Targeted advertising on social media or radio can be successful in recruiting men and can yield large numbers of responses. Having advertisements run consecutively for weeks seemed to result in cumulative benefits.
- General practitioner (GP) surgeries and word of mouth were good for recruiting a more balanced sample in terms of gender, ethnicity, and socioeconomic status (SES) but overall yielded lower numbers of participants. However, these methods were likely impacted by the COVID-19 pandemic and may be more effective with an increased investment of time or money in future trials.
- Offer technical support for online surveys and intervention use, ideally in different forms such as through pictorial step-by-step guides or through phone or email to ensure recruitment and engagement are inclusive.

Follow-up

- Offline follow-up options, such as phone calls and postal surveys, are more resource intensive but can increase follow-up rates.
- SMS text messaging services can be a relatively low-cost and low-effort way of boosting follow-up rates.

Advertising and incentives

- Avoid overly formal language, which may alienate participants, and use pictures.
- Highlight benefits to participants other than financial incentives (e.g., support for alcohol reduction).
- Tailor advertising strategies to ensure the right balance of incentivization across different platforms. For example, if advertising on social media or where barriers to sign up are low, mentioning incentives could result in motivated individuals falsifying information. However, where there are more barriers to sign up, for example, through a radio advertisement where participants must find the study link independently, it may be necessary to highlight incentives more explicitly.

Participant deception

- Be aware of different types of fraud and the best ways to detect them, and continuously monitor data as strategies are likely to evolve in response to checks and barriers introduced. These may include address checks, phone calls, or requiring participants to submit ID.
- When creating online surveys, researchers should use fraud detection software if it is offered (e.g., CAPTCHAS [Completely Automated Public Turing test to tell Computers and Humans Apart]) and check licenses to see if additional fraud detection software is available.
- Include attention-check questions where participants are asked to give stable information at different points in a survey or where participants are asked to select a particular response option.
- Ensure costing is included for the data monitoring resources required.

Contamination

- Consider the inclusion of sensitivity analyses, such as instrumental variable analysis, to capture the extent of contamination in remote randomized controlled trials.

Discussion

Summary of Findings

In this remote RCT, the 3 main participant recruitment methods were through advertisements on social media (2483/5602, 44.32%), the NHS website (1961/5602, 35.01%), and through radio or newspapers (745/5602, 13.3%), with all other methods of recruitment accounting for 7.37% (413/5602) of the sample. More participants who were women, White, and from more advantaged backgrounds responded to the initial recruitment. Targeted approaches through social media and radio advertising were successful in recruiting men but less successful in appealing to a more diverse demographic in terms of ethnicity and SES. The most effective methods for recruiting more balanced samples (health care providers and word of mouth) were often responsible for a relatively small proportion of the overall sample, suggesting greater investment in these methods could be a positive strategy in future trials. The costs associated with different recruitment methods varied. There was an increase in cost per participant when recruiting participants who were men, from ethnic minorities, and from more disadvantaged backgrounds across all recruitment methods.

There was evidence that the sequential approach taken to 6-month follow-up was successful, with 79.58% (4458/5602) follow-up rates at 6-months. Most participants responded following automated emails and substantial financial incentives, including an additional incentive to respond to the primary outcome within the first 24 hours, but each additional stage of follow-up resulted in an additional 2% to 3% of the sample following up. The advantage of the sequential approach is also evidenced by the greater follow-up rate (4458/5602, 79.58%) at 6-month follow-up when this process was followed, relative to the follow-up rates at 1- and 3-months (3685/5602, 65.78% and 3574/5602, 63.8%, respectively), where only email or SMS
text reminders were sent and less financial incentive was offered. However, each of the offline stages of follow-up was considerably more resource intensive than email reminders, so this is a practical consideration to be made at the costing stage. It would be of great interest to compare, across trials, the sociodemographic characteristics associated with the sample captured at each stage of follow-up. For example, it may be possible that offline stages of follow-up may be effective in retaining less digitally literate or less engaged participants.

Implications

When making methodological decisions about remote RCTs, there is rarely a right answer that is applicable to every study or circumstance. It is important to be aware of balancing forces, which often pull in different directions. For example, when considering advertising, it is important to balance making the study appealing to the target market with not making the study so appealing that it yields a high rate of participants who sign up with false information or who respond multiple times to gain financial reimbursement. There is a similar trade-off when considering processes aimed at reducing participant deception in the data. It is important that processes that aim to ensure participants are real and eligible do not add postrandomization bias to remote RCTs by removing “real” participants in potentially nonrandom ways. Part of navigating this balance is to plan carefully and tailor decisions to individual circumstances, as well as to monitor and learn from decisions made throughout a trial.

Previous Research

The findings of this study are in line with other studies that have focused on methodological issues in remote studies and RCTs [11,13]. The recruitment strategy undertaken was informed by a previous smoking cessation trial, which recommended using a range of sources but also monitoring the success of strategies throughout to recruit a large, diverse sample. We have reported on the success of each strategy here to inform the planning of future trials. An additional potential strategy that we did not use here to improve ethnic diversity in trial participation is geotargeting of social media advertisements in geographic areas with an ethnically diverse population [22]. The multistage follow-up strategy and stepped approach to incentives (eg, an additional £12 if completed within 24 hours at 6 months) undertaken throughout the IDEAS trial were also informed by previous research [11]. The need to have ongoing strategies to detect participant deception in web-based studies and trials is also supported in other studies, and other strategies recommended beyond those we used are to check participant IDs during onboarding and undertake IP address checks [13].

Limitations

This study offers valuable insights for researchers conducting web-based or remote RCTs, but it is not without limitations. The cost per participant is calculated for different sociodemographic groups to demonstrate the relative increase in costs required to recruit a balanced sample. However, this stepped increase in costs is conflated by narrowing the focus to smaller groups in the population. For example, we would expect that each participant from ethnic minority groups would cost more than each participant overall when simply dividing the cost by the number of participants, because there are proportionately fewer of them. Regardless, our estimates of comparative costs for different demographic groups across different recruitment methods may help other researchers who are planning future trials. Furthermore, this study does not consider costs related to setting up the trial, developing automation, designing materials for data collection and recruitment, and engaging with stakeholders to promote recruitment. These are additional upfront and ongoing costs that should be considered when costing RCTs. There are also 2 limitations related to the generalizability of these findings. Due to the very small numbers of some ethnic minorities, ethnicity was treated as White versus ethnic minority. Grouping all ethnic minority participants together in this way does not allow examination of different methods of recruitment for attracting different ethnic minorities. Furthermore, the Drink Less app is currently only available to those with an iOS device, and as such, iOS device ownership was an entry requirement for the trial. There are some sociodemographic differences in iPhone ownership: relative to Android devices, iPhone owners are younger, more likely to be women [23], and have higher average incomes [24].

Conclusion

Most participants in this remote RCT were recruited through advertisements on social media (2483/5602, 44.32%), the NHS website (1961/5602, 35.01%), and through radio or newspapers (745/5602, 13.3%). Most recruitment methods oversampled participants who were more advantaged, women, and White. Targeted approaches through social media and radio advertising were successful in recruiting men but less successful in appealing to a more diverse demographic in terms of ethnicity and SES. There was evidence that the sequential approach taken to 6-month follow-up was successful, with 79.58% (4458/5602) follow-up rates at 6 months. This study offers recommendations for achieving balance in methodological challenges when conducting remote RCTs. Recruitment methods should be broad and targeted to achieve sociodemographic diversity. Automated emails with substantial financial incentives can achieve excellent follow-up rates of approximately 70%, but sequential offline follow-up can further boost retention by nearly 10% overall. SMS text messages can be a low-cost, low-effort way to improve follow-up rates. An important and broader takeaway is the importance of continuously monitoring, identifying, reacting to, and documenting new methodological challenges as they appear over the course of a trial. This is necessary not only to improve individual trials but also because pooling shared experiential learning can help research teams who are planning future trials.
Acknowledgments
This study is funded by the National Institute for Health and Care Research (NIHR; Public Health Research Programme, #127651). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. MO’s salary is funded by NIHR (NIHR127651) and Medical Research Council (MRC; MR/W026430/1). CG is funded by NIHR (NIHR302923).

Data Availability
The data sets generated during and analyzed during this study are available on Open Science Framework [25].

Conflicts of Interest
MH, GL, LD, MF, and SM declare no conflicts of interest. JB has received unrestricted funding related to smoking cessation research and sits on the scientific advisory board for the Smoke Free app. CG and MO have done paid consultancy work for the behavior change and lifestyle organization “One Year No Beer,” providing fact-checking for blog posts.

References

https://formative.jmir.org/2024/1/e51839 JMIR Form Res 2024 | vol. 8 | e51839 | p.37 (page number not for citation purposes)

©Melissa Oldham, Larisa Dinu, Gemma Loebenberg, Matt Field, Matthew Hickman, Susan Michie, Jamie Brown, Claire Garnett. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.01.2024. 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.
Remote Delivery of the Cuidándome Telehealth Intervention for Self-Management of Depression and Anxiety Among Latina Immigrant Women: Randomized Controlled Trial

Carmen Alvarez1, MSN, PhD; Subhash Aryal1, MS, PhD; Elizabeth Vrany2, PhD; Maria Jose Sanchez R3, MD, MPH; Rosalphie Quiles4, PhD; Lia Escobar-Acosta5, BS; Felicia Hill-Briggs2†, PhD

1School of Nursing, University of Pennsylvania, Philadelphia, PA, United States
2Institute of Health System Science, Feinstein Institutes for Medical Research, Department of Medicine, Zucker School of Medicine, New York, NY, United States
3Department of Health Promotion, College of Public Health, University of Nebraska Medical Center, Omaha, NE, United States
4Division of Geriatric Medicine and Gerontology, Johns Hopkins University School of Medicine, Baltimore, MD, United States
5Johns Hopkins School of Nursing, Baltimore, MD, United States
†deceased

Corresponding Author:
Carmen Alvarez, MSN, PhD
School of Nursing
University of Pennsylvania
418 Curie Boulevard
Philadelphia, PA, 19104
United States
Phone: 1 2158980715
Email: alcarmen@nursing.upenn.edu

Abstract

Background: Growing evidence suggests that Latina immigrant survivors of adverse childhood experiences (ACEs) are at increased risk for developing and remaining with either depression or anxiety or both symptoms. This study examined the feasibility and acceptability of a telehealth intervention—Cuidándome (quee-DAN-doh-meh, “taking care of myself”). Cuidándome is a 10-week, patient-centered, trauma-informed intervention delivered by a trained facilitator that promotes self-management of depression and anxiety symptoms through improved problem-solving skills and strategies.

Objective: The aim of this study was to examine the feasibility and acceptability of Cuidándome delivered remotely (via Zoom) with Latina immigrant ACE survivors with either depression or anxiety or both symptoms. We also estimated the effect sizes associated with the intervention on decreasing depression and anxiety symptoms and improving social problem-solving styles.

Methods: We evaluated Cuidándome using a randomized controlled trial design. Latina immigrants (N=47) who had experienced at least 1 ACE and had at least mild depression or anxiety symptoms were randomized to Cuidándome or a comparison group delivered by trained facilitators. We assessed for changes in depression and anxiety symptoms as well as social problem-solving styles at baseline, post intervention, and 3- and 6-month follow-up.

Results: Analyses indicated significant decreases over time within both Cuidándome and comparison groups for depression and anxiety symptoms and maladaptive problem-solving. The intervention effect was largest for anxiety; at 6-month follow-up, Cuidándome participants had significantly lower anxiety scores than the comparison group. In addition, we observed a greater average point reduction in depression symptoms at 6 months among Cuidándome participants (5.7 points) than in the comparison group (3.7 points).

Conclusions: A mental health program delivered via Zoom by a trained facilitator was feasible and acceptable to Latina immigrant women and can be beneficial for reducing anxiety and depression symptoms. More research is needed to assess the effectiveness of Cuidándome among a powered sample size of Latina immigrants.

Trial Registration: ISRCTN Registry ISRCTN16668518; https://www.isrctn.com/ISRCTN16668518

doi:10.2196/52969
KEYWORDS
Latinas immigrant; mental health; depression; anxiety; problem-solving; intervention study; trauma-informed; depressive; Latinx; Latin; Latino; Latina; Hispanic; Spanish; immigrant; immigrants; survivor; child; children; childhood; trauma; traumatic; adverse; telehealth; telemedicine; eHealth; digital health; feasibility; acceptability; randomized; controlled trial; controlled trials; mobile phone

Introduction

Background

Latinas immigrant survivors of adverse childhood experiences (ACEs) are at increased risk of poor long-term health outcomes, as mental health disorders often go untreated in this population [1]. ACEs are a spectrum of adversities that occur in 18 years and younger of age and include physical and sexual abuse, family dysfunction such as living with an adult with mental illness, experiencing or witnessing community violence (eg, stabbing and shooting), and experiencing or witnessing violence perpetrated by an organized group (eg, gang violence and police or military brutality) [2]. These types of experiences are established risk factors for anxiety and depressive disorders in adulthood [1,3]. The United States has seen historic levels of immigrants from Central America—many fleeing from different types of adversity, trauma (eg, natural disasters, pervasive community, and political violence), and limited opportunity for socioeconomic advancement. Growing evidence suggests that foreign-born Latinos, particularly those from countries with high risks of community and political violence, experience high rates of early childhood adversity that are associated with poorer mental health outcomes [3]. When compared to the general US population and Latino immigrant men, Latina immigrants report significantly higher rates of multiple types of ACEs [1].

Despite the high burden of adversity and depression and anxiety symptoms, multiple barriers impede Latina immigrants’ access to mental health services. System-level barriers such as lack of health insurance and lack of language-concordant services are common barriers to accessing mental health services [4]. Despite the growth of the Latino population in the United States, there has been a decline in mental health services offered in Spanish [5]. In addition, while evidence-based psychological treatments are recommended for the general population, they are becoming less available in primary care settings [7]. Implementation of evidence-based psychological treatment services in Latino-serving health care settings also remains a challenge. These limitations make it difficult for Latino immigrants who prefer psychotherapy to pharmacotherapy to access the mental health care they need [8,9]. Studies targeting Latinos in primary care settings have often used licensed personnel as interventionists; however, the sustainability of providing such services in low-resource settings is questionable. To address these barriers, mental health experts recommend expanding access to behavioral health services by providing them outside of specialized settings, using telehealth services [10], and rigorously training and supervising unlicensed personnel (such as community health workers) to deliver high-quality services [11].

Teledotherapy and the use of paraprofessionals both show promise in increasing acceptability and engagement in the treatment of Latina immigrants experiencing depression. Among Latinos, adherence to teledtherapy sessions was higher (>80%) compared to in-person sessions (42%-80%) [12,13]. Despite the methodological limitations, paraprofessional-led interventions have demonstrated improvements in depression symptoms among Latinos [14,15]. Given the shortage of behavioral health professionals, delivering mental health care through paraprofessionals or community health workers is a promising strategy for increasing access for underserved populations [11,16].

Community health workers are also attuned to the needs of the communities they serve and have feasible solutions. Community health workers who serve Latino populations acknowledge (1) the need for more mental health services, (2) the training for community health workers to better meet this need, and (3) the use of teleservices to make care more accessible [17]. Community health workers have also proposed group support for addressing mental health needs. Indeed, the group format helps to reduce feelings of isolation and shame as participants hear from others who have similar life experiences with trauma and depression and anxiety symptoms [18]. Further, group support maximizes the community health workers’ reach as multiple individuals can be served and supported by each other.

Problem-solving is an established evidence-based approach for managing depression and anxiety symptoms. Social problem-solving refers to the cognitive behavioral process used to cope with life stressors [19]. According to problem-solving theory, coping with stressors involves two independent components: (1) problem orientation and (2) problem-solving style [20]. Problem orientation refers to one’s general cognition and attitudes when faced with a problem; this process is also framed by past experiences and self-appraisal about problem-solving ability. Problem-solving style refers to cognitive behavioral activities people use to cope with or manage stressful situations and include rational problem-solving (RPS; systematic and deliberate application of problem-solving skills), impulsive-careless style (ICS; impulsive approaches to problems), and avoidance style (AS; procrastination and avoiding addressing the problem) [21,22]. Effective social problem-solving involves identifying barriers to practicing recommended behaviors and brainstorming strategies to overcome barriers [23].

Among Latina immigrant women, ACE survivors had lower self-confidence in stress management compared to women who did not report ACEs [3]. In addition, experiencing more types of adversity was negatively associated with overall social problem-solving skills and positively associated with negative problem orientation (NPO) and AS [3]. Understanding and overcoming barriers through problem-solving underscores the importance of trauma-informed care, in which trauma survivors are supported in understanding how childhood adversities
contribute to mental and physical health and reducing negative self-evaluations that impact problem-solving styles [24]. To date, the most widely used and evaluated psychological intervention among Latino immigrants is cognitive behavioral therapy, and the established benefit to Latina immigrants is based on 3 randomized controlled trials with limited generalizability [25]. Randomized controlled trials testing problem-solving therapy for decreasing depression symptoms among Latina immigrants showed clinically significant reductions in symptoms up to a year postintervention when compared to pharmacotherapy [26,27]. In summary, there remains a need to expand the portfolio of effective mental health interventions to maximize reach and enhance responsiveness to diverse needs among Latina immigrants.

This Study
Given the lack of mental health services for Latina immigrants and the evidence for problem-solving and trauma-informed care, we developed Cuidándome (quee-DAN-doh-meh, “taking care of myself”). Cuidándome is a 10-week, culturally appropriate, trauma-informed, group-based intervention delivered once a week by a trained facilitator that promotes self-management of depression and anxiety symptoms through improved problem-solving skills and strategies. Multiple strategies were used to provide a trauma-informed intervention, including training the research team in therapeutic communication, screening and education about ACEs and their impact on health, and creating a safe and trusting environment for participants to work through their barriers for implementing useful strategies for depression and anxiety symptom management. Details of the adaptation and development process for Cuidándome are documented elsewhere [28]. The aims of this study were to (1) examine the feasibility and acceptability of Cuidándome delivered remotely (via Zoom; Zoom Video Communications) with Latina immigrant ACE survivors with either depression or anxiety or both symptoms and (2) estimate the effect sizes associated with the Cuidándome intervention on decreasing depression and anxiety symptoms and improving social problem-solving styles. We hypothesized that compared to the comparison group, the intervention group would report lower depression and anxiety symptoms, higher positive problem orientation (PPO) and RPS, and lower NPO, AS, and ICS at postintervention and at 3- and 6-month follow-up.

Methods
Recruitment
We recruited participants over 2 weeks in July 2021. Both active and passive strategies were used to recruit participants. Actively, we developed a database of Latina immigrants with a prior study [3] and contacted these women to assess for eligibility and participate in this study if interested. We also shared the study flyer with community health workers in the area, who distributed the flyer within their networks, including placing study flyers inside bags of food that were being donated during a food drive. Our passive strategies included posting flyers at laundromats and grocery stores in Latino-concentrated neighborhoods. Women who were interested in participating in the study texted or called the research phone. A bilingual research assistant obtained informed consent and established eligibility over the phone for all women verbalizing interest in participating in the study.

Establishing eligibility included the completion of a baseline study questionnaire (including demographic information and assessments for depression and anxiety symptoms) to verify eligibility for the study. Eligibility criteria included (1) being ≥18 years, (2) foreign-born (or born on the island of Puerto Rico), (3) self-identify as a Latina, (4) self-report of ≥1 ACE, (5) ability to understand and speak Spanish, and (6) have a score of ≥5 on the Patient Health Questionnaire-8 (PHQ-8)—an assessment for depression symptoms [29] or ≥5 on the Generalized Anxiety Disorder-7 (GAD-7)—an assessment for anxiety symptoms [30]. We excluded women currently enrolled in another study about mental health (to limit potential confounding or carryover effects), and we excluded women who reported being pregnant (given that pregnancy can contribute to depression symptoms). Figure 1 displays the CONSORT (Consolidated Standards of Reporting Trials) diagram, participant enrollment, and retention (Multimedia Appendix 1).
**Ethical Considerations**

All study procedures were approved by the Johns Hopkins University Ethics Review Board (IRB00287200). Oral consent was obtained from all participants in Spanish over the phone by the bilingual research assistants (native proficiency). To secure and protect all participant information, all data were collected and directly entered into REDCap (Research Electronic Data Capture) hosted at Johns Hopkins University. Only select research team members could access these data. All data were deidentified prior to export to SPSS (version 28; IBM Corp) for data analysis. Given the cost associated with data use, we compensated our participants up to US $190 for study participation (US $15 per session attended) and completion of all follow-up study questionnaires.

**Procedures**

Women who provided consent were found eligible for the study (based on the baseline study questionnaire), and women who agreed to enroll in the study were randomized to receive either Cuidándome or educational content from a health promotion manual designed in Mexico [31]. Randomization was stratified based on ACE score so that one group would not have more people with higher average ACEs than the other. After completion of the baseline questionnaire and randomization, participants were told when their group sessions would begin, and participants were mailed the corresponding workbook for
their group assignment. Based on input from community partners and the significant use of mobile phones and apps in the study population, we did not make computer or internet access a requirement for participation. For those who were not familiar with Zoom, a brief orientation was scheduled to explain how to use Zoom. All participants were encouraged to join the group sessions when they were scheduled; if a participant could not attend the group session, a make-up session would be scheduled with the participant where the facilitator would review content from the missed week prior to the next group session. This progression was particularly important because the sessions were designed to build on each other.

Study questionnaires were completed at baseline (T0) as part of the eligibility assessment and enrollment process, within 1 month post intervention (T1), and again at 3 months (T2) and 6 months (T3) post intervention. Trained bilingual research assistants who were not involved in the intervention delivery administered the study questionnaire (see “Study Questionnaire” section for descriptions of items and measures) to participants via phone and entered responses into a secure REDCap database. Based on our experience and prior evidence, many low-income Latino immigrants rely on their smartphones for internet access, particularly if they do not subscribe to broadband services [32]. Our retention efforts included 3 weekly reminders via SMS text message for joining the Zoom sessions, mailing participants a Cuidándome bookmark and a booklet of poems.

Intervention
Table 1 provides the content overview for the intervention and comparison groups. To summarize, Cuidándome facilitates the learning and practice of systematic problem-solving through identification of the problem, generation of potential solutions, selection of the best solution, and implementation of the identified plan. Given participants’ history of trauma, the intervention sessions start with content about mental health and how ACEs, as well as other types of adversity, can contribute to mental health symptoms and conditions in adulthood. The remainder of the sessions guides participants through 5 evidence-based self-management strategies for managing depression and anxiety symptoms and identifying solutions for the barriers (life activities and stressors) that get in the way of practicing the recommended strategies. The weekly sessions lasted approximately 1 hour. In addition to the facilitator, all participants had the Cuidándome workbook that provided structured templates for guiding participants through the session activities. The first session included a discussion about ground rules, including the importance of confidentiality and not sharing comments made within the group with people outside of the group. During each session, the facilitator encouraged group discussion and shared reflections and strategies for overcoming challenges. The first 2 modules focus on psychoeducation and allow for discussion throughout. In the remaining modules, group learning through participant discussion is the priority; therefore, the facilitator presents the activity, guides participants through the exercises, and encourages discussion using vignettes and the workbook.
Table 1. Brief description of modules for Cuidánde and comparison program.

<table>
<thead>
<tr>
<th>Cuidánde</th>
<th>Comparison (health promotion group)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1A: ACEs⁴, depression, anxiety, and PTSD⁵</strong></td>
<td></td>
</tr>
<tr>
<td>• ACEs and their associations with health</td>
<td>1: Physical activity</td>
</tr>
<tr>
<td>• Signs and symptoms of depression, anxiety, and PTSD</td>
<td>• What is physical activity?</td>
</tr>
<tr>
<td>• Mental health stigma</td>
<td>• Benefits of physical activity</td>
</tr>
<tr>
<td></td>
<td>• Suggestions for remaining physically active</td>
</tr>
<tr>
<td><strong>1B: Mental health and self-management strategies</strong></td>
<td></td>
</tr>
<tr>
<td>• Signs and symptoms (continued)</td>
<td>2A: Healthy eating</td>
</tr>
<tr>
<td>• Review and discussion of personal profile</td>
<td>• Review and discussion of food groups</td>
</tr>
<tr>
<td>• Self-management strategies for depression and anxiety</td>
<td>• Discussion of portion sizes</td>
</tr>
<tr>
<td></td>
<td>• Recommendations for healthy eating</td>
</tr>
<tr>
<td><strong>2: Overview of problem-solving</strong></td>
<td>2B: Healthy eating</td>
</tr>
<tr>
<td>• Identify behaviors for self-management of mental health</td>
<td>• Recommendations for health eating</td>
</tr>
<tr>
<td>• Identify barriers to self-management</td>
<td>• My BMI</td>
</tr>
<tr>
<td>• Understand the steps in problem-solving approach</td>
<td></td>
</tr>
<tr>
<td><strong>3: Taking control of stress and emotions (problem orientation)</strong></td>
<td></td>
</tr>
<tr>
<td>• Understand negative versus positive problem orientation and its impact on problem-solving</td>
<td>3: Mental health</td>
</tr>
<tr>
<td>• Understand the relationship between emotions and behavior</td>
<td>• What is mental health?</td>
</tr>
<tr>
<td></td>
<td>• Why is it important?</td>
</tr>
<tr>
<td></td>
<td>• What can influence your mental health?</td>
</tr>
<tr>
<td><strong>4: What makes a problem a problem? (problem identification)</strong></td>
<td></td>
</tr>
<tr>
<td>• Identify external and individual barriers to self-management</td>
<td>4: Substance misuse</td>
</tr>
<tr>
<td>• Demonstrate knowledge of the problem-solving process</td>
<td>• Addiction prevention</td>
</tr>
<tr>
<td></td>
<td>• Assessing alcohol consumption</td>
</tr>
<tr>
<td></td>
<td>• Smoking cessation</td>
</tr>
<tr>
<td><strong>5: Know thyself: set goals that fit your life (generating alternative solutions)</strong></td>
<td></td>
</tr>
<tr>
<td>• Understand the importance of identifying problems for appropriate goal setting</td>
<td>5: Chronic diseases</td>
</tr>
<tr>
<td>• Demonstrate an understanding of effective goal setting</td>
<td>• Prediabetes and diabetes</td>
</tr>
<tr>
<td></td>
<td>• Hypertension</td>
</tr>
<tr>
<td></td>
<td>• Hyperlipidemia</td>
</tr>
<tr>
<td><strong>6: Different ways to reach health goals: knowing yourself</strong></td>
<td></td>
</tr>
<tr>
<td>• Understand the importance of exploring multiple options for problem-solving</td>
<td>6: Cancer screenings</td>
</tr>
<tr>
<td></td>
<td>• Breast cancer risk factors and screening</td>
</tr>
<tr>
<td></td>
<td>• Cervical cancer screening</td>
</tr>
<tr>
<td><strong>7: That sounds good but does it work for me?</strong></td>
<td></td>
</tr>
<tr>
<td>• Understand one’s own values and priorities in decision-making and problem-solving</td>
<td>7: Osteoporosis</td>
</tr>
<tr>
<td>• Demonstrate understanding of the 4 problem-solving styles and the impact on problem-solving</td>
<td>• What is osteoporosis?</td>
</tr>
<tr>
<td>• Identify rational problem-solving as the effective approach for solving problems</td>
<td>• Prevention of osteoporosis</td>
</tr>
<tr>
<td><strong>8: Take action and know the signs</strong></td>
<td></td>
</tr>
<tr>
<td>• Acquire skills for attempting alternative solutions for solving problems</td>
<td>8: Respiratory illnesses</td>
</tr>
<tr>
<td>• Demonstrate awareness of signs that a solution is not working</td>
<td>• Prevention and control of communicable respiratory infections</td>
</tr>
<tr>
<td></td>
<td>• Prevention and control of noncommunicable respiratory illnesses</td>
</tr>
<tr>
<td><strong>9: Putting it all together</strong></td>
<td></td>
</tr>
<tr>
<td>• Demonstrate mastery of the rational problem-solving approach</td>
<td>9: Review</td>
</tr>
<tr>
<td>• Articulate the problem-solving approach for the management of mood</td>
<td>• What did you learn?</td>
</tr>
<tr>
<td></td>
<td>• What has helped you?</td>
</tr>
<tr>
<td></td>
<td>• What will you continue to do for your well-being?</td>
</tr>
</tbody>
</table>

⁴ACE: adverse childhood experience.
⁵PTSD: posttraumatic stress disorder.

**Comparison**

Given our focus on Latina immigrant women with either depressive or anxiety or both symptoms, we opted to offer the comparison group some generic health education content rather than be waitlisted. The content for the comparison group came from a family health promotion manual from Instituto Mexicano del Seguro Social—the Mexican Institute for Social Security [31]. We selected this manual given our focus on a Spanish-speaking, immigrant Latina population. The content from Instituto Mexicano del Seguro Social was already in Spanish and culturally appropriate—particularly the nutrition...
content that referenced traditional foods and diets common to our participants. The comparison content included 1 session about mental health conditions in general, which did not overlap with the more detailed Cuidándome training. Delivery of the comparison content mirrored the format of the intervention group: 1-hour weekly sessions delivered via Zoom.

**Group Facilitators and Intervention Fidelity**

The intervention facilitator is a bilingual, Latina immigrant with a bachelor degree, who was trained in problem-solving therapy by our expert clinical psychologist and who received ongoing support and guidance. The facilitator of the comparison group was an experienced registered nurse with a master in health education and expertise (over 15 years of experience) in facilitating health promotion groups for Latina immigrant women. Aside from training on human participants, the nurse for the comparison group did not receive any specific training but was oriented to the purpose of the comparison group and provided with the corresponding workbook content. All Zoom sessions were audio recorded and reviewed after the sessions by the principal investigator to assess for client-centeredness (eg, showing empathy and encouraging autonomy) in both groups and to determine if the Cuidándome facilitator followed the facilitator script and demonstrated a problem-solving approach (eg, guiding participants to targets for change and focusing on positive action).

**Study Questionnaire**

**Demographic Characteristics**

The study questionnaire included questions about demographic characteristics: age, relationship status, children, nativity, length of time in the United States, educational attainment, and employment status. Items about demographic characteristics were only administered at baseline.

**Adverse Childhood Experiences**

The Adverse Childhood Experiences-International Questionnaire was used, at baseline only, to assess for occurrence (eg, “Did you live with a household member who was a problem drinker or alcoholic, or misused street or prescription drugs?” “Yes” or “No”) and frequency (eg, “Did a parent, guardian or other household member hit or cut you with an object, such as a stick (or cane), bottle, club, knife, whip etc.” “Many times,” “A few times,” “Once,” or “Never”) of different types of adversities that occurred in the age of 18 years and younger [33]. In addition to items that inquired about the traditional ACEs (eg, physical and emotional neglect), the Adverse Childhood Experiences-International Questionnaire also assesses for the types of adversity such as bullying and experiencing or witnessing community violence. Items about child marriage were not included in our assessment because we have not identified this experience as a significant part of our population’s history. We dichotomized item responses based on the presence or nonzero frequency of an experience (yes=1 and no=0) and summed all dichotomized item responses for a total score; higher scores indicated experiencing more types of adversity. This tool has been validated with Latina immigrants [3,34].

**Primary Outcome Variables**

**Depression**

We used the PHQ-8 to assess the frequency (0=not at all to 3=nearly every day) of depression symptoms during the last 2 weeks [29]. Item responses are summed for a total score (range 0-24), with higher scores indicating greater severity of symptoms. The PHQ-8 has been validated among Latina immigrants and demonstrated good reliability with our sample ($\alpha=.83$).

**Anxiety**

We used the GAD-7 to assess the frequency (0=not at all to 3=nearly every day) of anxiety symptoms during the last 2 weeks [30]. Item responses are summed for a total score (range 0-21), with higher scores indicating greater severity of symptoms. The GAD-7 has also been validated among Latina immigrants and demonstrated good reliability with our sample ($\alpha=.76$).

**Social Problem-Solving**

We assessed social problem-solving styles using the Social Problem-Solving Inventory-Revised (SPSI-R) [20]. The items assess attitude toward challenges as well as one’s tendencies and approach for managing stressors in everyday life. Items present different styles of thinking and reactions to scenarios to which participants report how accurately the statement reflects their attitudes or behaviors to challenges (0=not at all true of me to 4=extremely true of me). The SPSI-R assesses for (1) problem orientation and (2) problem-solving style. Problem orientation refers to one’s disposition and attitude toward a problem. People with a PPO perceive problems as solvable challenges and are optimistic and confident in their ability to manage the problem; higher scores on the PPO subscale indicate greater confidence and optimism for solving problems. People with an NPO tend to perceive problems as a threat and are less confident in their ability to address the problem; higher scores on the NPO indicate less confidence in their ability to address problems.

Three problem-solving styles are assessed with the SPSI-R: RPS, ICS, and AS. When faced with challenges, people who practice RPS tend to think through multiple solutions and intentionally implement the optimal approach; higher scores on the RPS subscale indicate higher levels of RPS. The ICS is used to describe the tendency to act on the first option that comes to mind rather than consider multiple solutions; higher scores on the ICS subscale indicate greater impulsivity when addressing problems. The AS describes the practice of procrastination or avoiding addressing a problem; higher scores on the AS subscale indicate greater tendency for practicing avoidance for addressing problems. Each subscale was added for a sum score. To obtain a total social problem-solving score, the subscales are calculated (maladaptive styles negatively impact the total score social problem-solving score) using the prescribed formula [20]. These assessments have been used among Latina immigrants and demonstrated good reliability ($\alpha=.74$).
Statistical Analysis

Descriptive statistics (frequencies, means, and SDs) for all participant demographics and outcome variables were calculated. We tracked the number of sessions completed for each participant as an indicator of acceptability and asked participants if and how the intervention helped them at the end of the 10 weeks. The proportion of interested participants who consented and were screened as eligible to be in the study was also computed to help inform feasibility. We conducted independent t tests to assess differences in outcome variables (depression, anxiety, and social problem–solving styles) between the intervention and comparison groups at baseline and the follow-up time points (data not shown). In addition, paired t tests were used to compare the differences in means for the outcome variables from postintervention to 3- and 6-month follow-up time points (data not shown). For our primary analyses, mixed between-within participants’ ANOVA was used to examine differences in outcome variables between the intervention and comparison group, over time, from baseline to 6-month postintervention. We calculated effect sizes (Cohen d: small <0.50, medium ≥0.50 to <.80, and large ≥0.80) using the difference in outcome means for the different groups divided by the pooled SDs.

Results

Participant Demographics, ACEs, and Retention

Our sample included 47 Latina immigrants at baseline (Table 2) and 41 participants at all follow-up assessments. There were no significant differences between the intervention and comparison group at baseline.
## Table 2. Participant characteristics\(^a\).

<table>
<thead>
<tr>
<th></th>
<th>Total sample(^a) (N=47)</th>
<th>Intervention group (n=23)</th>
<th>Comparison group (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>35.72 (8.4)</td>
<td>36.78 (9.2)</td>
<td>34.7 (7.6)</td>
</tr>
<tr>
<td><strong>Relationship status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>8 (17)</td>
<td>3 (6)</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Married</td>
<td>16 (34)</td>
<td>8 (17)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>Living together (not married)</td>
<td>20 (43)</td>
<td>10 (21)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Living apart (not married)</td>
<td>2 (4)</td>
<td>1 (2.1)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Divorced</td>
<td>1 (2)</td>
<td>1 (2)</td>
<td>N/A(^b)</td>
</tr>
<tr>
<td><strong>Children, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>4 (8)</td>
<td>3 (6)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>1-3</td>
<td>35 (75)</td>
<td>19 (41)</td>
<td>16 (34)</td>
</tr>
<tr>
<td>4 or more</td>
<td>8 (17)</td>
<td>1 (2)</td>
<td>7 (15)</td>
</tr>
<tr>
<td><strong>Nativity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico</td>
<td>11 (23)</td>
<td>5 (10)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>El Salvador</td>
<td>9 (19)</td>
<td>5 (11)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>Guatemala</td>
<td>5 (10)</td>
<td>2 (4)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Honduras</td>
<td>12 (26)</td>
<td>4 (9)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>Other (Caribbean and South America)</td>
<td>10 (21)</td>
<td>7 (15)</td>
<td>3 (6)</td>
</tr>
<tr>
<td><strong>Length of time in the United States, mean (SD)</strong></td>
<td>10.48 (6.4)</td>
<td>11.3 (6.2)</td>
<td>9.7 (6.6)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school or less</td>
<td>9 (19)</td>
<td>3 (6)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Some high school education</td>
<td>9 (19)</td>
<td>7 (15)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>High school graduate or more</td>
<td>29 (62)</td>
<td>13 (28)</td>
<td>16 (34)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not employed</td>
<td>26 (55)</td>
<td>16 (34)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Employed full-time</td>
<td>12 (26)</td>
<td>3 (6)</td>
<td>9 (19)</td>
</tr>
<tr>
<td>Employed part-time</td>
<td>9 (19)</td>
<td>4 (9)</td>
<td>5 (11)</td>
</tr>
<tr>
<td><strong>Adverse childhood experiences, mean (SD)</strong></td>
<td>11.26 (4.8)</td>
<td>11.1 (5)</td>
<td>11.4 (4.6)</td>
</tr>
<tr>
<td><strong>Depression symptoms, baseline, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None (&lt;5)</td>
<td>6 (13)</td>
<td>3 (6)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Mild (5-9)</td>
<td>18 (38)</td>
<td>8 (17)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Moderate depression (10-14)</td>
<td>12 (26)</td>
<td>6 (13)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Major depression, moderately severe (15-19)</td>
<td>6 (13)</td>
<td>4 (9)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Major depression, severe (20-24)</td>
<td>5 (11)</td>
<td>2 (4)</td>
<td>3 (6)</td>
</tr>
<tr>
<td><strong>Anxiety symptoms, baseline, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None (&lt;5)</td>
<td>2 (4)</td>
<td>N/A</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Mild (5-10)</td>
<td>23 (49)</td>
<td>12 (26)</td>
<td>11 (23)</td>
</tr>
<tr>
<td>Moderate (10-14)</td>
<td>14 (30)</td>
<td>9 (19)</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Severe (15-21)</td>
<td>8 (17)</td>
<td>2 (4)</td>
<td>6 (13)</td>
</tr>
</tbody>
</table>

\(^a\)Baseline sample.
\(^b\)N/A: not applicable.
The most common ACEs included community violence (n=39, 83%), witnessing violence in the home (n=39, 83%), emotional abuse (n=36, 77%), physical abuse (n=36, 77%), and being bullied (n=35, 75%). In addition, 55% (n=26) of the sample reported some form of sexual abuse (unwanted sex, fondling, and attempted sex).

Figure 1 displays participant enrollment, retention, and adherence to the group sessions. Of the participants we assessed for eligibility, most (n=38, 63%) were women who contacted the research team indicating their interest to participate. The other participants were selected from the database for a previous study. Attrition was low, with 6 women discontinuing participation primarily due to work schedules. All Cuidándome participants (n=20, 100%) completed at least 9 of the 10 total sessions, and 76% (n=16) of the comparison group completed 9 of the total 10 sessions.

### Depression and Anxiety and Social Problem-Solving Overview

In Tables 3 and 4, we present the mean scores for depression, anxiety, and social problem-solving styles by study group (intervention and comparison) and time (baseline, postintervention, and 3- and 6-month follow-up). In Table 5, we compared for differences of change in scores between the intervention and comparison groups for depression, anxiety, and social problem-solving styles (interaction effect); compared the change in depression, anxiety, and social problem-solving styles over time within the groups (time main effect); and compared the 2 programs for changing depression, anxiety, and social problem-solving styles (intervention main effect).

#### Table 3. Group mean scores for depression, anxiety, and social problem-solving styles at baseline, postintervention, and 3- and 6-month follow-up.

<table>
<thead>
<tr>
<th>Time point</th>
<th>Depression(^a)</th>
<th>Anxiety(^b)</th>
<th>Social problem–solving styles(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention, mean (SD)</td>
<td>Comparison, mean (SD)</td>
<td>Effect size(^d)</td>
</tr>
<tr>
<td>Baseline</td>
<td>10.75 (5.19)</td>
<td>10.19 (6.12)</td>
<td>N/A(^e)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>4.15 (3.04)</td>
<td>5.95 (4.26)</td>
<td>0.48</td>
</tr>
<tr>
<td>3-month follow-up</td>
<td>6.10 (4.66)</td>
<td>7.80 (6.16)</td>
<td>0.31</td>
</tr>
<tr>
<td>6-month follow-up</td>
<td>5.05 (2.95)</td>
<td>6.47 (4.58)</td>
<td>0.37</td>
</tr>
</tbody>
</table>

\(^a\)Patient Health Questionnaire-8 [29].
\(^b\)Generalized Anxiety Disorder-7 [30].
\(^c\)Social Problem-Solving Inventory-Revised [20].
\(^d\)Cohen \(d\): difference in outcome means for the different groups divided by the pooled SDs.
\(^e\)N/A: not applicable.

#### Table 4. Group mean scores for social problem-solving styles at baseline, postintervention, and 3- and 6-month follow-up.

<table>
<thead>
<tr>
<th>Time point</th>
<th>Negative problem orientation(^a)</th>
<th>Avoidance style(^b)</th>
<th>Impulsive-careless style(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention, mean (SD)</td>
<td>Comparison, mean (SD)</td>
<td>Effect size(^b)</td>
</tr>
<tr>
<td>Baseline</td>
<td>9.3 (4.29)</td>
<td>10.62 (4.90)</td>
<td>N/A(^c)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>6.1 (2.4)</td>
<td>6.42 (4.13)</td>
<td>0.10</td>
</tr>
<tr>
<td>3-month follow-up</td>
<td>5.45 (3.15)</td>
<td>7.66 (4.82)</td>
<td>0.54</td>
</tr>
<tr>
<td>6-month follow-up</td>
<td>4.85 (3.20)</td>
<td>6.42 (4.72)</td>
<td>0.38</td>
</tr>
</tbody>
</table>

\(^a\)Social Problem-Solving Inventory-Revised [20].
\(^b\)Cohen \(d\): Difference in outcome means for the different groups divided by the pooled SDs.
\(^c\)N/A: not applicable.
Table 5. Intervention and time effects on depression, anxiety, and social problem-solving styles.

<table>
<thead>
<tr>
<th></th>
<th>Depression</th>
<th>Anxiety</th>
<th>Social problem-solving</th>
<th>Negative problem orientation</th>
<th>Avoidance style</th>
<th>Impulsive-careless style</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention by time interaction</strong></td>
<td>Wilks $\Lambda$</td>
<td>$F$ test (df)</td>
<td>Wilks $\Lambda$</td>
<td>$F$ test (df)</td>
<td>Wilks $\Lambda$</td>
<td>$F$ test (df)</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.99</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td><strong>Time main effect</strong></td>
<td>0.53</td>
<td>11.1 (3, 37)</td>
<td>0.51</td>
<td>11.9 (3, 37)</td>
<td>0.44</td>
<td>16.0 (3, 37)</td>
</tr>
<tr>
<td><strong>Intervention main effect</strong></td>
<td>N/A ^d</td>
<td>0.91 (1, 39)</td>
<td>N/A</td>
<td>4.3 (1, 39)</td>
<td>N/A</td>
<td>1.6 (1, 39)</td>
</tr>
</tbody>
</table>

^a$P<.001$.
^b$P<.005$.
^c$P=.02$.
^dN/A: not applicable.

**Depression**

Based on the PHQ-8, depression levels decreased from baseline to postintervention for both groups and remained below baseline at 3 and 6 months (Table 3). There were small effect sizes (Cohen $d$) at each time point (postintervention (T1): $d=0.48$; 3-month follow-up (T2): $d=0.31$; and 6-month follow-up (T3): $d=0.37$; Table 3). Depression symptoms significantly increased over time (main effect) for both intervention and comparison groups (Wilks $\Lambda=0.53$; $F_{3,37}=11.1$; $P<.001$; Table 5). However, when comparing the 2 groups, the change in depression symptoms over time was not significant (intervention by time interaction in Table 5). There was also no significant difference between the 2 programs in reducing depression symptoms (intervention main effect in Table 5); specifically, at each time point, there was no significant difference in depression symptoms between the groups.

**Anxiety**

Based on the GAD-7, anxiety levels also decreased from baseline to post intervention and remained below baseline through 6 months for both groups (Table 3). We estimated small effect sizes (Cohen $d$) at each time point (T1: $d=0.30$ and T2: $d=0.36$) and medium (T3: $d=0.65$) effect sizes for reduced anxiety symptoms (Table 3). The reduction in anxiety symptoms over time was significant, with both groups showing a reduction in symptoms across the follow-up time points (Wilks $\Lambda=0.51$; $F_{3,37}=11.9$; $P<.001$; Table 5). There was also a significant difference in the reduction of symptoms between the 2 groups, where Cuidámdone was demonstrated to be more effective than the comparison program for reducing anxiety symptoms ($F_{1,39}=4.3$; $P<.001$).

**Social Problem-Solving Styles**

Overall social problem-solving increased from baseline to all 3 time points. Similar to depression and anxiety symptoms, the increase in social problem-solving over time was significant with both groups showing improvement (Wilks $\Lambda=0.51$; $F_{3,37}=11.9$; $P<.001$). When comparing the 2 groups, the increase in social problem-solving scores over time was not significant (intervention by time interaction in Table 5). There was also no significant difference between the 2 programs for increasing social problem-solving scores (intervention main effect in Table 5). The ANOVA analyses indicated that neither program had a significant effect on PPO or RPS (data not shown).

**Negative Problem Orientation**

NPO decreased from baseline to all 3 time points (Table 4). The reduction in NPO over time was significant for both groups (Wilks $\Lambda=0.44$; $F_{3,37}=16$; $P<.001$; Table 5). Although there was change over time for both groups, this change did not differ by group (intervention by time interaction), and there was no significant difference in the effect of the 2 programs for reducing NPO (intervention main effect). However, the difference in NPO at 3-month follow-up between the 2 groups almost reached significance (Cuidámdone: mean 5.45, SD 3.15 vs comparison: mean 7.66, SD 4.82; $P=.09$), and we estimated small (T1: $d=0.10$) and medium (T2: $d=0.54$ and T3: $d=0.38$) effect sizes (Table 4).

**Avoidance Style**

ASs decreased from baseline to all 3 time points (Table 4). Similar to NPO, the reduction in AS over time was significant for both groups (Wilks $\Lambda=0.68$; $F_{3,37}=5.8$; $P<.005$; Table 5). There was no difference in change over time between the 2 groups, and there was no significant difference in the effect of the 2 programs for reducing ASs; however, the difference in mean scores postintervention (Cuidámdone: mean 2.1, SD 2.17 vs comparison: mean 4.42, SD 4.81; $P=.06$) and at 6-month follow-up (Cuidámdone: mean 1.65, SD 2.5 vs comparison: mean 4.0, SD 4.80; $P=.06$) approached significance, where Cuidámdone participants reported lower AS at these time points (Table 4). We estimated medium effect sizes (T1: $d=0.62$, T3: $d=0.61$) for the intervention.

**Impulsive-Careless Style**

ICS decreased from baseline to all follow-up time points (Table 4). ICS decreased over time for both groups (Wilks $\Lambda=0.75$; $F_{3,37}=4.2$; $P=.02$); however, the change in scores did not differ significantly between the groups, and neither program was more effective at reducing ICS scores (Table 5). Nonetheless, we...
observed lower mean scores for Cuidándome participants compared to the comparison group that approached significance at 3-month follow-up (Cuidándome: mean 4.55, SD 4.08 vs comparison: mean 7.1, SD 4.54; \( P=0.07 \); Table 4). We calculated small (T1: \( d=0.14 \)) and medium (T2: \( d=0.59 \) and T3: \( d=0.40 \)) effect sizes for the intervention.

Discussion

Principal Findings

This study is one of the first to assess the feasibility and acceptability of a trauma-informed, problem-solving–based, self-management program delivered remotely for Latina immigrant ACE survivors with at least mild depression and anxiety symptoms. Our rapid recruitment (completed in 2 weeks), high attendance, and retention indicated that participants desired the program and found it acceptable. Based on the PHQ-8 means, participants in both groups were experiencing moderate levels of depression symptoms at baseline. Depression symptoms significantly decreased for both groups, with no significance in change between the intervention and comparison group. However, we observed lower depression scores among Cuidándome participants compared to the comparison group suggesting Cuidándome participants experienced fewer days with depression symptoms. Although both groups experienced improvements, Cuidándome participants on average reported a greater reduction in depression symptoms (5.7 points) compared to the comparison group (3.7-point reduction). For anxiety symptoms, Cuidándome was significantly more effective at reducing anxiety symptoms. On average, Cuidándome participants reported minimal to no anxiety symptoms at all follow-up time points compared to comparison group participants who on average reported mild symptoms. The effect sizes for depression and anxiety were small to medium, further supporting the beneficial impact of Cuidándome for these symptoms. Further study with a powered sample is needed to rigorously test the effectiveness of Cuidándome in this Latina immigrant population.

Contrary to our hypothesis, we did not see improvements in PPO or RPS—the components of social problem-solving that we anticipated Cuidándome would increase. Instead, we found that among Cuidándome participants, we observed lower scores for NPO and maladaptive problem-solving styles (avoidance and impulsive-careless). The goal setting and learning the problem-solving steps may have helped Cuidándome participants feel more inspired and empowered to address daily life challenges in order to pursue their goals. When discussing the benefits of Cuidándome, our participants shared that Cuidándome provided them with the steps for “how” to achieve their goals; this may have helped women have a more positive outlook on addressing challenges [28]. Women also shared that they felt a greater sense of confidence managing daily challenges and thinking through options before reacting to a situation. Regarding the lack of findings with the positive subscales, although we did not include assessments of social desirability, social desirability may have influenced participant responses and minimized the scales’ sensitivity to change. Further, we were not powered to identify statistically significant changes with any of our outcomes.

Other studies that have examined the effectiveness of problem-solving therapy for depression among Latinos have also identified improvements in depression symptoms [27,35]. However, this study is the first to show promising findings on anxiety as well as social problem-solving styles, which are the potential mechanisms of action for improving mental health outcomes.

We unexpectedly observed significant reductions in depression and anxiety symptoms in both groups. On review of the session recordings, we learned that the nurse facilitator for this group used both goal-setting and problem-solving (particularly brainstorming solutions) strategies in her sessions—particularly for the nutrition and physical activity sessions. At the end of these sessions, participants were encouraged to set a goal based on the session topic, and they discussed strategies for achieving those goals. Participants also exchanged contact information with fellow participants in the chat feature of Zoom. We did not assess social support, but it is possible that through these sessions, participants were able to expand their social network and increase social support, which is associated with lower depression and anxiety symptoms [36]. In addition, given that there were 2 sessions about physical activity and nutrition, comparison group participants may have increased their physical activity; indeed, increased physical activity is associated with a reduction in depression symptoms among Latina women [37] and other populations [38]. Finally, for all participants, we made ourselves available to connect them with community resources. Participants often called for information about where they could find health care services, work, and food. In our future work, we will assess whether change in social needs is associated with improved mental health.

Based on prior work, we surmise that the trauma-informed content and care from our research team were important contributors to retention. During the development phase of Cuidándome, the review of ACEs and their association with mental health conditions were the most time-consuming sessions because of participant engagement. Similar to findings by Kaltman et al [39] who also examined the feasibility and acceptability of an in-person, trauma-informed intervention, Cuidándome participants had positive reviews about the discussions on trauma, and they found it validating to learn that their current depression and anxiety symptoms could be related to early life adversities. Participants who were mothers felt inspired to engage with their children in a more positive way to not perpetuate the cycle of ACEs. Qualitative analysis of participant discussions during these sessions may provide more insight into participant responses to the trauma-informed content.

The success of this feasibility study may also be attributed to the intervention being offered remotely. Multiple structural (eg, documentation status) [40] and system-level barriers (eg, accessibility, health insurance, and language concordant services) make mental health services and care inaccessible for Latina immigrant women [41]. Cuidándome eliminated several of these macrolevel barriers—there was no need for participants...
to present themselves in any establishment with government-issued identification in order to obtain services, participants did not have to travel to a physical location, health insurance was not required, and the program was offered in Spanish. Using telehealth and trained personnel eliminated barriers that prevent marginalized groups from accessing a program that may be beneficial for mental health. More research with stakeholders is needed to determine how programs such as Cuidándome can be made more accessible and sustainable in community-based settings.

Limitations
We acknowledge several limitations with this study. First, we sought to establish acceptability and feasibility and did not calculate a sample size a priori. Our relatively small sample size may explain the few statistically significant findings between Cuidándome and the comparison program. In addition, our sample represented women primarily in urban and suburban settings with access to broadband services. A larger sample size that includes some geographic diversity may yield more generalizable findings.

Despite the limitations, this work contributes to the body of literature highlighting specific useful strategies (telehealth and nonlicensed personnel) that can be used to expand access to mental health services for populations socially at risk and underserved populations. Nonlicensed personnel such as community health workers have successfully delivered mental health services in low-resource settings [16]. This work aligns with other studies demonstrating the acceptability and effectiveness of training nonlicensed personnel to deliver mental health interventions [39,42] as well as the use of a web-based platform for administering these programs.

Conclusions
Our findings indicate that the Cuidándome intervention can improve depression and anxiety symptoms among Latina immigrant ACE survivors. Further, Cuidándome may also be beneficial for decreasing maladaptive behaviors (avoidance and impulsivity) associated with depression and anxiety symptoms. As the Latina immigrant population continues to grow, so should community-based mental health resources. More methodologically rigorous study of Cuidándome is needed; however, this study shows the promise of an intervention that leverages nonlicensed personnel and uses a web-based platform to increase the availability of a beneficial mental health program.

Acknowledgments
FH-B developed the program on which Cuidándome is based. This study was supported by the Robert Wood Johnson Foundation Harold Amos Faculty Development Program.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to ethical considerations but are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
CONSORT e-HEALTH (V.1.6.1) checklist. [PDF File (Adobe PDF File), 410 KB - formativ8i1e52969_app1.pdf ]

References


33. Adverse Childhood Experiences International Questionnaire (ACE-IQ). World Health Organization. URL: https://www.who.int/publications/m/item/adverse-childhood-experiences-international-questionnaire-(ace-iq) [accessed 2020-08-01]

Abbreviations

ACE: adverse childhood experience
AS: avoidance style
CONSORT: Consolidated Standards of Reporting Trials
GAD-7: Generalized Anxiety Disorder-7
ICS: impulsive-careless style
NPO: negative problem orientation
PHQ-8: Patient Health Questionnaire-8
PPO: positive problem orientation
REDCap: Research Electronic Data Capture
RPS: rational problem-solving
SPSI-R: Social Problem-Solving Inventory-Revised
Edited by A Mavragani; submitted 20.09.23; peer-reviewed by F Cross; comments to author 19.10.23; revised version received 16.11.23; accepted 29.11.23; published 08.01.24.

Please cite as:
Remote Delivery of the Cuidándote Telehealth Intervention for Self-Management of Depression and Anxiety Among Latina Immigrant Women: Randomized Controlled Trial
 JMIR Form Res 2024;8:e52969
URL: https://formative.jmir.org/2024/1/e52969
doi:10.2196/52969
PMID:38190239

©Carmen Alvarez, Subhash Aryal, Elizabeth Vrany, Maria Jose Sanchez R, Rosalphie Quiles, Lia Escobar-Acosta, Felicia Hill-Briggs. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.
Feasibility and Acceptability of a Mobile Technology Intervention to Support Postabortion Care After Surgical Abortion (the FACTS Study Phase 3): Mixed Methods Prospective Pilot Study

Roopan Gill 1,2,3, BSc, MD; Gina Ogilvie 2,4, MD, PHD; Wendy V Norman 2,4, MD, CCFP, DTM&H, MHSC; Brian Fitzsimmons 2,3, MD; Ciana Maher 2, BA; Regina Renner 2,3, MD, MPH

1 Department of Obstetrics and Gynecology, University of Toronto, Toronto, ON, Canada
2 Women's Health Research Institute, University of British Columbia, Vancouver, BC, Canada
3 Department of Obstetrics and Gynecology, University of British Columbia, Vancouver, BC, Canada
4 Department of Family Practice, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Roopan Gill, BSc, MD
Department of Obstetrics and Gynecology
University of Toronto
123 Edward Street
Suite 1200
Toronto, ON, M5G1E2
Canada
Phone: 1 4169782216
Email: roopan.gill@gmail.com

Abstract

Background: In Canada, 1 in 3 women and people of gestational age undergo an abortion in their lifetime. Despite the liberal legal context, barriers continue to exist for women and people who can become pregnant to access this service.

Objective: This study aims to (1) conduct a pilot study to demonstrate the feasibility and acceptability of myPostCare to support follow-up care after a procedural abortion; (2) use the findings to understand whether myPostCare has the potential to improve contraceptive behavior and knowledge, emotional well-being, and sexual health knowledge; and (3) develop a better understanding of how innovative mobile solutions can support integrative health programs in British Columbia with the goal of expanding to other sites across Canada.

Methods: People of gestational age (aged 14–45 y) who underwent a procedural abortion were recruited from 2 urban abortion facilities in British Columbia. The participants completed a baseline quantitative survey and were provided access to myPostCare for up to 30 days. A follow-up quantitative survey was sent via email on day 30. Qualitative interviews were conducted to explore user satisfaction and usability of myPostCare. Responses to the survey questions were summarized using descriptive statistics, and the system usability scale (SUS) was scored according to the instructions. A secure analytics platform was implemented to obtain data on the overall use of the website by users. Qualitative analysis was conducted with NVivo using a thematic analysis approach. This study was approved by the Women’s and Children’s Research Ethics Board.

Results: Overall, 62 participants were recruited (average age 30 y); 40% (25/62) of the participants completed the exit surveys, and 24% (6/25) consented to participate in the semistructured interviews; 40 participants had undergone an immediate postabortion intrauterine device (IUD) insertion, and 22 did not have an IUD inserted. Participants were satisfied with myPostCare. The SUS average score was 81.5 (SD 9.7; median 82.5, IQR 77.5–87.5), indicating high usability of the tool. Overall, 88% (22/25) of the participants changed their contraceptive method to an IUD. Web-based analytics demonstrated that there were 61 unique visitors to the site, and the top pages visited were Postprocedure Care, Emotional Well-Being, and Contraception Explorer. The longest time spent on the website was 56 minutes. The overall email open rate was 80%, with a click rate of 36%.

Conclusions: This study demonstrates that communities and individuals are important collaborators in developing a mobile innovation that facilitates access to high-quality patient-centered abortion care. Through the cocreation process, a digital platform such as myPostCare highlighted a gap in abortion care in Canada, particularly around follow-up support after a procedural abortion.
Furthermore, there is a great deal of inconsistency in the type of support and information available to women and people who can become pregnant after an abortion. The New England Journal of Medicine published a special report on Telehealth in the United States, highlighting its utility and future. In 2016, Kaiser Permanente of Northern California reported that its virtual (email, telephone, and video) communications had exceeded in-person visits [11]. Similarly, research supports the safe and effective use of telehealth for the provision of medication abortion care globally [12-14].

Three trials of mHealth interventions have aimed to study the role of mobile interventions in increasing the use of contraception [15-18]. Mobile for Reproductive Health and Mobile Alliance for Maternal Action have used best practices from health communication programs to systematically develop family planning text messages [18]. Furthermore, Smith et al [13,14] explored women’s needs in Cambodia to develop a mobile phone–based intervention to support postabortion family planning, specifically contraceptive adherence. In the United States, research on the acceptability and feasibility of remote technologies for follow-up after medication abortion suggested that women prefer either a telephone call or a text message over a clinic visit [19]. Most recently, researchers from University of San Francisco’s Program in Women-Centered Contraception developed a tablet-based contraceptive decision support tool for women [20]. This study used a multiphase approach that incorporated the end user throughout the entire design of the project. The tool has been designed in collaboration with key stakeholders and designers from Bedsider [21]. Using an iterative process informed by patient and provider input throughout, this family planning innovation demonstrated that including users in development led to a more patient-centered innovation [22]. Despite the development and implementation of these mHealth innovations for family planning, research is limited in understanding the follow-up needs of women and people who can become pregnant and undergo an abortion, and how they would perceive a tool to support them and to engage them as active participants in the design process.

Given the existing evidence in support of mHealth for family planning innovations, we aimed to determine whether a mobile technology intervention would be acceptable and feasible for women and people who can become pregnant to support follow-up care after first or second trimester procedural abortion. We developed a 3-phased study based on human-centered design and the “Development-evaluation-implementation” process from the Medical Research Council Frameworks for Complex Interventions [23] rooted in 2 theories: Technology Acceptance Model and Theory of Reasoned Action [24,25] phases 1 and 2 have been published previously [26,27]. This study was a prospective pilot that aimed to determine whether the intervention was satisfactory, acceptable, and usable for women and people who can become pregnant to support them in follow-up after a procedural abortion. Ultimately, this study is...
the first to use mHealth and human-centered design in Canada as a novel approach to support follow-up care for women and people who can become pregnant and undergo procedural abortion.

**Methods**

**Participants**

Participants were recruited from 2 publicly funded abortion clinics in British Columbia, Canada. The eligibility criteria were as follows: (1) consent to undergo a first or second trimester procedural abortion, (2) ability to read and write English, (3) ability to participate in study procedures, and (4) aged ≥14 years. Participants were excluded if they were (1) attending the clinics because of fetal anomaly or miscarriage, (2) undergoing medication abortion, (3) in a situation where it may be dangerous to use a mobile intervention, and (4) unable to provide consent to participate. To elicit whether a woman was in a dangerous situation, counselors asked the patients as part of routine care if they felt safe in their current relationships. In cases where a risk is identified, counselors provided resources and would refer to the appropriate provider or service.

**Study Design**

The overall study design is a mixed methods user-centered design approach with 3 phases based on the “Development-evaluation-implementation” process from the Medical Research Council Frameworks for Complex Medical Interventions [28]. This is the final phase of the 3-phase study, with the findings from phases 1 and 2 already published [26,27]. Phase 3 is a prospective pilot mixed methods study conducted in 2 urban clinics in Vancouver, British Columbia, between March and June 2018 to test the acceptability and feasibility of myPostCare when implemented as part of clinical care. This study was approved by the Children’s and Women’s Research Ethics Board (H18-00036).

Eligible participants were screened by a primary investigator. They were then introduced to the study and consented under the supervision of the research coordinator. Participants consented to be contacted for a qualitative interview at 4 weeks. A baseline questionnaire that was adapted from validated survey tools was filled out to collect demographic information, contraception history, and levels of perceived well-being and distress in the past 2 months before the abortion [29-31]. The Arizona Integrative Outcomes Score was used and is a validated 1-item visual analog tool that allows self-rated global assessment of spiritual, social, mental, emotional, and physical well-being over the past 24 hours and 1 month [32].

The participants were registered on the website at the end of each recruitment day. Participants received 7 automatic email notifications that were timed with what would be expected after the procedure and prompted them to the website over the course of 30 days. At the end of 30 days, participants received a link to their email to complete a questionnaire adapted from the validated questionnaires [29-31,33,34]. This questionnaire specifically included questions about satisfaction with myPostCare, a system usability scale (SUS) comprising 10 questions, and an evaluation of the impact of various aspects of myPostCare including emotional well-being, contraceptive behavior, immediate postprocedural care, and sexual health. Data analytics were collected using a secure data analytic platform housed at the BC Children’s Hospital. Participants were compensated for their participation.

Participants who consented to the qualitative part of the study were contacted and invited to participate in semistructured interviews to explore their engagement with the mobile tool. This included system usability, experience of receiving email notifications, emotional well-being, contraceptive decision-making, immediate postprocedural concerns, and questions about sexual health. Questions explored experience with receiving timed email messages, feedback on the content of the notifications themselves, if they found the notifications helpful and why, did they follow the recommendations of the notifications, and did they visit the website after being prompted by the notifications. Participants received additional compensation for their participation in the interview.

**Data Analysis**

Descriptive analysis of each variable from the quantitative surveys and secure data analytic platforms was reported as mean (SD) or median for continuous variables and count (percentage) for categorical variables. All statistical analyses were performed in R (R Foundation for Statistical Computing). Using Piwik, a secure web analytics through the BC Children’s Hospital Research Institute, specific user engagement data were gathered from February 20 to May 2, 2018. The semistructured interview transcripts were uploaded to NVivo 11 (Lumivero) and read by 2 researchers. Inductive analysis was performed to identify emerging themes that were further refined through collaborative analysis with the first author and coinvestigator [35]. The highlighted text was coded into nodes representing similar or repeated ideas. Some text was coded to >1 node, reflecting the number of ideas presented. The nodes were categorized into specific themes, forming a thematic map that was later discussed with the research team. To enhance the validity of the findings, a triangulation approach was used. This involved cross-referencing data from the quantitative survey and the subsequent 2 phases of this study.

**Ethical Considerations**

This study received ethics approval by the Children’s and Women’s Research Ethics Board (H18-00036). Informed consent was obtained from all the participants included in this study. The study data were anonymized and deidentified. All data were stored in an encrypted file only accessible to the research team involved in the analysis of the study. Compensation was not provided to those who had completed the survey. A CAD $25 (US $18.38) honorarium was provided to those who completed an interview.

**Results**

**Participant Characteristics**

Participants were recruited from 2 abortion clinics in Vancouver, British Columbia. A total of 62 participants were recruited and completed the baseline survey. Of the 62 participants recruited, 25 (40%) women and people who can become pregnant.
responded to the follow-up survey. We investigated whether systematic differences existed between women who responded and those who did not. Table S1 in Multimedia Appendix 1 provides a summary of the demographic information from the baseline survey and a comparison between responders and nonresponders. There were no substantial differences between these 2 groups for any of the variables listed, although there was a nonsignificant trend for the responders to have a lower Arizona Integrative Outcomes Score. These results were not statistically significant \((P<.05)\). All the participants identified as ciswomen.

For the qualitative interviews, of the 25 participants who completed the exit survey, 6 (24%) consented to participate in semistructured individual interviews. These were conducted via telephone.

**Quantitative**

**Change in Contraceptive Method**

Most of the respondents (22/25, 88%) indicated that they had changed their contraceptive method to an intrauterine device (IUD) at the time of their abortion, and 21 (95%) of the 25 respondents indicated that they had changed to a Mirena, whereas 1 (4%) of the 25 respondents indicated changing from a copper to Mirena. The contraceptive method of choice was not influenced by the website; however, the website and email notifications helped reassure participants about the signs, symptoms, and effectiveness of the IUD.

**System Usability Scale**

The SUS comprised 10 questions [36]. The average SUS was 81.5 (SD 9.7), and the median was 82.5 (IQR 77.5-87.5), which revealed that 75% (19/25) of the respondents indicated an SUS score >77, which is a very high score.

**Satisfaction**

Most of the respondents were satisfied with the website. Figure 1 graphically displays these results as percentages.

**Qualitative**

**Overview**

Qualitative analysis of the interviews was completed using thematic analysis including both inductive and deductive themes. Nine key themes were identified and are listed in Textbox 1.
Textbox 1. Key themes.

<table>
<thead>
<tr>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ease of use</td>
</tr>
<tr>
<td>2. Usefulness of myPostCare</td>
</tr>
<tr>
<td>3. Website</td>
</tr>
<tr>
<td>4. Frequency of use</td>
</tr>
<tr>
<td>5. Time spent on the website</td>
</tr>
<tr>
<td>6. Suggestions for improvement</td>
</tr>
<tr>
<td>7. Recommend to friend</td>
</tr>
<tr>
<td>8. Privacy and security</td>
</tr>
<tr>
<td>9. Design features</td>
</tr>
<tr>
<td>10. Overall impressions of myPostCare</td>
</tr>
</tbody>
</table>

**Ease of Use**
Overall, there was unanimous agreement that myPostCare was easy to use with an organized and easy-to-navigate design. One participant highlighted the following:

> I thought it was very easy to use, which I really liked. I felt the information was well laid out with the menu sidebar on the side. The writing was easy to interpret and was not overly scientific. It was easy to navigate throughout the whole website. It didn’t feel like I was reading a research article. It was nicely spaced out and got to the point very quickly. [participant 3]

Furthermore, the language was accessible and user friendly. The participants felt that the drop-down features were very effective. We added this feature after the usability testing from phase 2, and therefore, it was consistent among our participants to hear that they appreciated this feature. The participants were highly satisfied with the ease of use of myPostCare. When asked about ease of use as it pertained to the information, 1 participant shared the following:

> The information there was superb. It was very user-friendly. Anyone could use that and get what they were looking for, no problem. [participant 5]

**Usefulness of myPostCare**
We asked about the overall usefulness of myPostCare by asking separately about the website and email notifications. Participants were satisfied with the overall frequency and timing of the email notifications:

> I like there was one email per week, it was not overwhelming. It gave you time to go back to the website in increments, not getting overwhelmed and not having it constantly on your mind, but it was a good refresher every week. This is what I needed. [participant 3]

> The timing was impeccable when you would get these e-mails and what you would be feeling. When they would come, they were right on point. I always felt like someone was at my fingertips if I needed help. [participant 5]

They found that the emails helped to navigate the recovery process from immediate signs and symptoms, emotional well-being, and contraception decision-making to general sexual health, such as a better understanding of their menstrual cycle. They also found that the emails helped them feel supported and not alone. This was an important point that resonated with all participants interviewed. A few participants stated the following:

> When I would get the email it would say, “Okay, now you might be going through this and this and this,” it gave me a moment to be like, “Right, I am. I might be going through this. I’m still having some symptoms. How am I actually feeling?” It was a reminder to check in with myself and also to think about how I may be experiencing symptoms at that time. [participant 1]

> It was nice to feel as though there was “someone” checking up on you even though it wasn’t a person. There was new content with each e-mail and helped to direct you to different stages of recovery process. I found that helpful. [participant 1]

The participants unanimously stated that myPostCare provided them the support that was needed at the right time. It was helpful for the resource to provide support over time and that it allowed them to navigate various aspects of their postcare journey. A few participants shared the following:

> I felt like I was cared for. It was amazing to get, “Hey, I hope you’re doing okay. Take care.” It just felt that someone was there for me and saying if you need to call or anything, you can at any time. [participant 2]

> I think this is a great resource. It was a really beneficial thing for me to have, for sure. [participant 6]

**Website**
Overall, all participants stated that they did not have a favorite page but that each category was helpful depending on the stage at which they were in the recovery process. Each participant mentioned that the Postprocedure care page and the Emotional Well-Being Support tool were the most effective:
The emotional well-being tool was helpful. I liked how each emotion had a little blurb about. I liked the meditation. [participant 2]

Talking about various emotions that occur was important because I found that one week I felt one way but then all of a sudden I would feel different. It was nice to go back to the website, have those feelings identified and made me feel normal. [participant 3]

Most found that the emails were well timed with the website, and providing links embedded within the emails to direct participants to the website was appreciated. Participants stated that they did not click on the emotion “Good” but did use the suggestions provided such as the meditation, journaling, and going for a walk:

I wrote an entire journal entry one day, and that was really good and definitely got some crying out while I was doing that, so I think the website prompted me to do that day, yes. [participant 3]

Just going through and trying to be at one with this, checking in, using the tools. There was good days, bad days. I have a wonderful program at work as well but I didn’t have to reach out to it because there was stuff here about meditation and making sure that I am looking after myself and doing something nice for myself. [participant 5]

In addition, a participant commented that the website had credible information, which helped to answer questions that she would have seen her physician about and, therefore, kept her out of the office or emergency room. When further asked if the website helped her understand when to seek hospital care, she mentioned that it was very clear. She specifically found the disclaimer useful in preventing women from misunderstanding the website as a substitute for clinical care:

It [myPostCare] kept my husband and I out of emergency rooms...Here we are, two weeks and three days, and all of a sudden there’s an email about UIUs being that you could have spotting for three to six months. I am like, “Okay, We are good.” Then the bleeding stopped. It was just very empowering to have that information. [participant 5]

Finally, some participants found the website useful to support them as they did not have anyone else to talk to about their abortion, and the website helped them not feel isolated:

I went every time I got an e-mail and then there were two times that I went on it by myself when I was feeling pretty emotional and the emotional support tool helped. [participant 2]

It was good for me because I didn’t tell anybody. I didn’t have anyone to talk to. [participant 4]

It [myPostCare] is so critical, and I hope it never goes away and that it’s there for as long as women need this procedure. I hope that this site is always there. It was truly instrumental to my whole well-being through this whole procedure, so I thank you. [participant 5]

Frequency of Use

Participants used the website on its own but also clicked on the links within the emails. Some participants saved all the emails so that they could return to them. Using the website for 1 month seemed to be sufficient for all participants. One participant stated that she had visited the website 10 to 15 times:

I went every time I got an e-mail and then there were two times that I went on it by myself when I was feeling pretty emotional and the emotional support tool helped. [participant 2]

Time Spent on Website

The time spent on the website, as expressed in the qualitative interviews, was consistent with the results from the web-based analytics presented in the Web-Based Secure Analytics section. Most participants stated that they were on the website for anywhere from 2 minutes to 1 hour:

I’d probably went all together 10 to 15 times. There was one time I was on it for probably an hour, but the other times, it was probably anywhere between two to five minutes. [participant 2]

I’d say around half an hour. [participant 3]

Maybe the first couple of weeks, I kind of looked at it. I looked through it for half an hour at dinner, 20 minutes, 15, half an hour. [participant 4]

...going to guess at least an hour going back through, making sure I didn’t miss anything going to the link, so at least an hour. [participant 5]

Suggestions for Improvement of myPostCare

There was a strong sentiment to include blogs and stories shared by women and people who can become pregnant and have undergone an abortion. This was available on the website, but these were found under the “Good” emotion. Many did not necessarily explore this section and mentioned that if they were feeling good, they did not necessarily explore the emotional support tool and would be keener on using the Contraception Explorer or Sexual Health pages. Furthermore, suggestions to add these on the main landing page or to have rotating articles that are specific to women and people who can become pregnant telling their stories would be very useful. This was highlighted by participants as a means of further enhancing the community feeling and not feeling alone in their experience. When further explored, this also highlighted that sharing stories was also a way to help destigmatize the experience that many women and people who can become pregnant and who participated in our interviews had internalized. The following excerpts highlight this theme:

I really wanted to hear someone’s story that was positive. I would have liked to listen to just having a couple of people’s stories and how it affected them just to compare myself to them. I don’t know. [participant 2]

Putting up videos or even having articles on different stories. [participant 2]
For me personally, if I’m feeling good or when I was starting to feel good about myself again, I wouldn’t have gone on the website to check that. [participant 2]

I think more testimonials and more quotes that you can use on that website from people who have been through the experience, the better because it gives validation for what women are going through and kind of makes us feel less alone. [participant 6]

The more testimonials and the more feedback you can get from women of all ages, all experiences, all the better. [participant 6]

**Recommend to a Friend**

All participants who were interviewed would recommend myPostCare to a friend. Some also suggested that this would be specifically good for friends who did not necessarily feel comfortable going to their physician or who did not have a family physician with whom they had a trusting relationship:

Yes, I think I totally would. I think it’d be very helpful to have. I don’t think it’s going to solve a friend’s problems or anything, but for me, it was helpful to have. [participant 2]

Yes, and it’s definitely one that I want to, like, if I ever know somebody that is going through that, I’m definitely going to recommend that to them. [participant 3]

I really would. This is a great resource for the person that doesn’t think that their issues always warrant a call to their doctor. [participant 6]

**Privacy and Security**

Our participants were satisfied with the level of privacy and security afforded by the emails and website. In particular, they noted that the emails were separate from the website, and some participants suggested that it would be essential to keep it this way when myPostCare would be made live. One participant stated the following:

It’s very discrete, and I liked that. The login is required to get on the website, so to me, it was certainty sufficient. My name is not all over the website, so even if I left it open, it is what it is, who knows what I was in. It’s not too specific so I was never worried if I had it open in public. [participant 5]

**Design Features**

All participants stated that the design was professional and the language was unbiased. Many participants commented that the design of the website and emails was calming and supportive. They also enjoyed the consistency between the website and email notifications. Words such as “clearly thought out,” “pleasant and cool pictures,” “nice blues and greys,” and “positive and well-crafted” came up frequently among our interview participants. One participant commented the following:

I really liked the resources simply because it went beyond just what we went through. Yes, I think that was one of my favorite or one of the things that when I got to, I was like, “Okay, there are crisis lines, and there are counselors.” Yes, of course, that’s what I expected to be on there. It went past that. It went to sexuality. It went to LGBTQ, or it went through different topics, so I feel like it was good education beyond what I just went through. [participant 2]

Yes, I was really happy I signed up for it and I was getting those e-mails weekly. I was able to access it, once again, read about different perspectives. I think overall, I don’t know if this is weird to say, but it was very calming. Approachable in a sense. It doesn’t hurt your eyes to stay on the website for a while. I really liked the colours. The layout was easy. [participant 6]

Other participants noted that the site was structurally thought through and that the design was relatable to them. One participant highlighted the following:

I thought it was clearly thought out and structurally too. The language is nicely worded and was very unbiased. [participant 2]

It was very nice and pleasant, the pictures were very cool. I liked the ranges of things that were on there, the whole thing about meditating and then also just needing actual straight up information was really helpful too. [participant 2]

Very soothing colours. The nice blues and then greys, yes. [participant 2]

I thought it was very easy to use. I felt that the information was well laid out with the menu sidebar. The images were quite big and spaced out so had to scroll quite a bit and not get through a lot of information. [participant 3]

There was no harsh colours. There was no in-your-face type of things that popped into the website. I liked there were no advertisements. I think the peaceful colourings, the “click this if you feel called to.” It’s nice to have that sense of well-being with positivity on a sensitive topic, it was well crafted. [participant 3]

**Overall Impressions of myPostCare**

Overall, the participants were satisfied with myPostCare. They felt supported by the resource. There were very strong sentiments that this went above and beyond what they had expected. One participant stated that she was surprised to have such a good experience with a website, as she had never had such an experience. Some of the participants used the site with their partners and appreciated the section that was specifically for partners. One participant stated that the emails and website helped to keep her out of the emergency department, as it highlighted the normalcy of postprocedural recovery. In general, all participants felt that it was a great experience to have this resource, and many participants expressed that this resource should be available as long as women need abortions:

I really liked the resources simply because it went beyond just what we went through. Yes, I think that was one of my favorite or one of the things that when I got to, I was like, “Okay, there are crisis lines, and there are counselors.” Yes, of course, that’s what I expected to be on there. It went past that. It went to sexuality. It went to LGBTQ, or it went through different topics, so I feel like it was good education beyond what I just went through. [participant 2]

Yes, I was really happy I signed up for it and I was getting those e-mails weekly. I was able to access it, once again, read about different perspectives. I think
there were some things that I felt like it was only me or it wasn’t normal, and then it would say something on the website that would make me feel better, more calm. [participant 2]

I would just grab my phone and then just go, look at the thing and, “Okay, this is normal to feel like this.” I don’t know if I had a favourite part, but I just found that everything was useful. [participant 4]

It was my other rock. My husband was my one rock, and the other one was this. It knew when things were going to happen, and when I was panicking about things, all of a sudden, there will be an e-mail. It was just perfect timing, and it was amazing. It truly was. I felt like I wasn’t alone. I went through every link. Even the links that were outside the website, I checked out every one of them. I read stories. It brought a sense of calm to me, I guess. It was truly, I never had such a good experience off of a website like this one. It was amazing. My husband went through everything. You would be panicking. I don’t know how many times we went back to this website to make sure that something that was going on wasn’t out of the ordinary, and of course, there would be, that it wasn’t out of the ordinary, so it was amazing. It truly was. [participant 5]

No, I think overall it was pretty straightforward. There wasn’t anything that I was surprised to see, and there wasn’t anything that I can remember that didn’t kind of fit in with what was expected through the e-mails. It all kind of made sense. [participant 6]

I think it was just a great experience to trial the website. I have my own personal reasons for my procedure and how I came about doing so, but I think it’s a great source for people that want to have that sense of community. I think it works really well for the specific areas that you’re trying to find more clarity. [participant 6]

Web-Based Secure Analytics

Table 1 presents the analytics results. Specifically, of the 62 participants, the number of unique visitors on the website was 61 (98%). Although only 25 participants completed the exit survey, all participants except 1 (98%) visited the website at least once. The number of returning visitors was 42. The average daily page views were 5; the total number of page views through the study period was 432; the highest number of hits at a single visit was 35; and the top 3 pages were Postprocedure Care, Emotional Well-Being, and Contraception Explorer. In total, 75% (47/62) of the participants were mobile users and 25% (16/62) were desktop visitors. The most popular contraceptive page visited was the IUD. The details of the number of page views throughout myPostCare are presented in Table 2.
Table 1. myPostCare web-based analytics for user engagement from February 20, 2018, to May 2, 2018 (N=62).

<table>
<thead>
<tr>
<th>Web-based analytics data</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique visitors to the website, n (%)</td>
<td>61 (98.4)</td>
</tr>
<tr>
<td>Average time spent on the website by visitors</td>
<td>1 min and 28 s</td>
</tr>
<tr>
<td>Longest visit on the website</td>
<td>35 hits</td>
</tr>
<tr>
<td>Total number of page views, n</td>
<td>432</td>
</tr>
<tr>
<td>Average daily page views, n</td>
<td>5</td>
</tr>
<tr>
<td>Participants who are mobile users, n (%)</td>
<td>47 (75)</td>
</tr>
<tr>
<td>Participants who are desktop visitors, n (%)</td>
<td>16 (25)</td>
</tr>
<tr>
<td>Participants who visited the Emotional Well-Being page, n (%)</td>
<td>28 (45)</td>
</tr>
<tr>
<td>Participants who visited the Contraception Explorer page, n (%)</td>
<td>27 (43)</td>
</tr>
<tr>
<td>Participants who visit the Postprocedure Care page, n (%)</td>
<td>46 (74)</td>
</tr>
<tr>
<td>Participants who visit the Sexual Health page, n (%)</td>
<td>21 (33)</td>
</tr>
<tr>
<td>Top 3 pages on the website</td>
<td>Postprocedure Care, Emotional Well-Being, and Contraception Explorer</td>
</tr>
</tbody>
</table>

Most popular contraceptives visited from the contraception tool in page views, n

- Hormonal IUD\(^a\) | 13 |
- Sterilization | 7 |
- Copper IUD | 3 |
- Vaginal ring | 3 |
- Fertility awareness | 2 |
- Patch | 2 |
- Abstinence | 1 |
- Depo shot | 1 |
- Female condom | 1 |
- Male condom | 1 |
- Withdrawal | 1 |

Visits to given feelings (good, okay, and not so good) from the Emotional Well-Being tool\(^b\)

- Okay | 18 |
- Good | 2 |
- Not so good | 5 |

Visits to given emotion from the Emotional Well-Being tool, n\(^b\)

- Grief | 6 |
- Relief | 6 |
- Supported | 4 |
- Sadness | 2 |
- Guilt | 3 |
- Regret | 1 |
- Shame | 1 |

\(^a\)IUD: intrauterine device.

\(^b\)Returning and 1-time visitors.
Table 2. Number of page views for myPostCare.

<table>
<thead>
<tr>
<th>myPostCare pages</th>
<th>Values, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postprocedure Care</td>
<td>46</td>
</tr>
<tr>
<td>Emotional Well-Being</td>
<td>28</td>
</tr>
<tr>
<td>Contraception Explorer</td>
<td>27</td>
</tr>
<tr>
<td>Postprocedure FAQs</td>
<td>25</td>
</tr>
<tr>
<td>Emotional Support Tool</td>
<td>21</td>
</tr>
<tr>
<td>Sexual Health</td>
<td>21</td>
</tr>
<tr>
<td>Abortion Myth or Fact Quiz</td>
<td>15</td>
</tr>
<tr>
<td>Resources</td>
<td>13</td>
</tr>
<tr>
<td>Menstrual Cycle 101</td>
<td>12</td>
</tr>
<tr>
<td>Meditation 101: Meditation for Beginners</td>
<td>10</td>
</tr>
<tr>
<td>Book an Appointment</td>
<td>6</td>
</tr>
<tr>
<td>Dealing with Difficult Feelings</td>
<td>6</td>
</tr>
<tr>
<td>About Us</td>
<td>5</td>
</tr>
</tbody>
</table>

Email Notifications Analytics

Among the 62 participants enrolled, 2 (3%) unsubscribed from email notifications after the “Welcome Message” on day 0, and 2 (3%) participants’ email address was not valid. The average open rate was 80%, and the click rate was 36%. The highest open and click rates were for Welcome Message at 73.1% and 31.3%, respectively. Interestingly, the desktop device was 57.4% and mobile was 42.6%, which is different from the device from which the website was viewed. The email open rates were higher than the click rates throughout. The open rates declined for both IUD and no IUD over time; however, they remained stable at an average of 53.7% and 53.8%, respectively. Figure 2 graphically represents data of the IUD versus no IUD streams open and click rates for given days.

Figure 2. Comparison of email open and click rates for intrauterine device (IUD) versus no IUD stream in percentage.

Discussion

Principal Findings

myPostCare is the first comprehensive web-based postabortion tool in Canada and has the potential to be integrated as part of family planning services. Integration of myPostCare into clinical practice provides an opportunity to consider a new approach to supplement follow-up care for abortion care specifically but women’s health generally. This study demonstrates the design and development of a comprehensive mobile intervention to facilitate care for women and people who can become pregnant and undergo a procedural abortion to support and normalize the emotional and physical aspects after abortion. We used a human-centered design methodology, an iterative development process that was informed by input from key stakeholders such as patients, family planning experts, and administrators involved in abortion care [22,37,38]. The results from the pilot evaluation of myPostCare demonstrated that it was feasible, acceptable, and satisfactory for women and people who can become pregnant.

Specifically, this 3-phase study demonstrates the importance of including the end users and key stakeholders in the design, development, and evaluation of a mobile intervention that services a population and health care issue that continue to be stigmatized. Formative research has provided important information regarding women’s interactions with technology, their needs and desires around follow-up and access to information, and feedback on design, which is essential for the success of myPostCare. A unique finding of this study that was
supported in the literature was the importance of including a component of emotional support as part of follow-up abortion care [39]. Furthermore, we learned that the success of myPostCare was not only owing to the interactive tools and information provided by the website but that the appropriately timed automatic email notifications that women received was an important aspect of their care throughout the 30 days after the procedure. An iterative design process was important to ensure that the research team was continually evaluating that myPostCare realized the needs of the target users.

We adopted a few theoretical frameworks, all of which use a comprehensive participatory approach to developing eHealth technologies. This was similarly performed by Gilbert et al [38] in the development of Get Checked Online, a web-based sexually transmitted infection testing resource. More specifically, integrating the Technology Acceptance Model and Theory of Reasoned action with the human-centered design methodology, we used a holistic approach to developing myPostCare. According to the Technology Acceptance Model, perceived ease of use and perceived usefulness of a system are the 2 predominant indicators of system adoption [27,32]. Participants in our study were accustomed to using some form of technology, either mobile phones or computers; did not require acquisition of new skills; and were keen on developing a technology-based tool to support follow-up care after an abortion. Importantly, myPostCare will not eliminate structural barriers to comprehensive abortion care, and although it may not directly affect health behavior and decision-making, it may assist in making the delivery of abortion care more efficient, convenient, patient centered, and accessible.

myPostCare is a unique addition to the literature because of its methodology and outcome. There is evidence to support eHealth technologies to improve health care; however, currently there is limited research on mobile interventions specifically to address postabortion care, although there are various interventions for contraception use. A randomized trial in Cambodia demonstrated that involving women in the design and testing of a mobile intervention to support postabortion contraception led to more women in the intervention group reporting use of effective contraception at 4 months; specifically, the use of long-acting contraceptives was higher in the intervention group at 4 and 12 months after the procedure [40]. Previous feasibility trials focused on usability and acceptability have highlighted the importance of conducting a pilot study first, which can then assist with the design of a larger randomized trial to measure effectiveness [41]. Finally, similar to studies on the development and testing of contraception tools, the integration of evaluation in real-time clinical care is essential to ascertain the barriers and challenges to implementation in the future [22].

The limitations of this study include overall generalizability to other populations, small convenience sample sizes for all 3 phases, loss to follow-up and low response rates in this challenging population, and recruitment bias. The sample size of 6 participants in the qualitative interviews was small, ideally requiring 20 participants to achieve meaningful saturation. Given that this study is an extension of 2 previous phases, researchers felt confident in the analysis being generalizable compared with the findings of the 2 previous phases and from previous studies highlighting the type of gaps that myPostCare fills as per the participants’ reflections. As it pertains to recruitment bias, those who consented to participate were likely individuals who are more engaged with technology, have higher socioeconomic demographics, and are more likely to be early adopters of a digital health intervention to support abortion care. In previous studies, this is referred to as a Digital Divide, which suggests that although many developers of technology-based health interventions are optimistic about their impact; this needs to be balanced by the fact that the pattern of adoption is along social gradients [38]. New technologies such as myPostCare may further reinforce these social divides. Furthermore, abortion continues to be a stigmatized issue, which can be a limitation for research, as this can be a sensitive topic for most and posed difficulties with recruitment and loss to follow-up in our study. We evolved throughout each phase of the study to consider the challenges faced with patient engagement. For instance, recruitment took longer than expected for the qualitative interviews. We assumed that lack of participant engagement may be associated with stigma about abortion. In addition, we recognized that conducting research immediately after the procedure might be a sensitive time for individuals. This will need to be taken into consideration for future studies, particularly when thinking about diversifying the participants recruited and obtaining robust response rates for analysis.

Balancing these limitations are the strengths of our study, including the successful development of human-centered design elements, wide stakeholder engagement, diverse expertise on the research team, a large proportion of our sample size that was from rural locations, rigorous research methodologies, iterative design process, and development of the first web-based postabortion tool in Canada.

Further research could involve evaluating the effectiveness of myPostCare.ca and the overall patient experience through a randomized controlled trial. In addition, as suggested in other web-based literature [38], a health equity impact assessment with expert consultation and literature review may also help identify ways in which myPostCare reinforces or alleviates health inequities in sexual health services.

**Conclusions**

myPostCare was found to be feasible and acceptable to women and people who can become pregnant to support follow-up care after a procedural abortion. There are obvious digital divides in health care specifically, as there are limited digital tools for women’s health in Canada. Thus, there is great potential for expansion of myPostCare. More specifically, since the introduction of Mifepristone in Canada, the first area of expansion will be for medication abortion. Generally, the expansion may then involve other aspects of women’s reproductive health.

We learned that key stakeholder engagement and understanding the organizational context are important. These factors are important for ongoing research initiatives and their implementation in clinical practice. Engaging stakeholders and potential users in a participatory process throughout the entire design and development of myPostCare was crucial to its
success. Applying an iterative design and evaluation process that was flexible and dynamic, considering the factors of implementation at the outset, keeping in mind how myPostCare could change health care delivery, and the use of a multidisciplinary team were all unique and important aspects. This study demonstrated that a technology-based intervention for postabortion care is feasible and acceptable. The success of myPostCare was based on the incorporation of a multidisciplinary team; participatory user-centered design process; robust stakeholder engagement; and the provision of nonjudgmental, nondirective, and medically accurate information. This study provides an example of the ongoing development of technology-based family planning services and is aligned with a larger gender-equitable, evidence-based programmatic agenda in Canada.

Acknowledgments
The authors thank the clinics that participated in this study and offered time.

Data Availability
Qualitative and quantitative data are available upon request; however, the data will be destroyed 10 years after its collection for this study.

Conflicts of Interest
WVN declares funding support for work contributing to this article for a Chair in Family Planning Public Health Research from the Canadian Institutes of Heath Research and The Public Health Agency of Canada (2014-2024) and for a Scholar award from the Michael Smith Foundation for Health Research (2012 - 2020).

Multimedia Appendix 1
Demographic data.
[PDF File (Adobe PDF File), 146 KB - formative_v8i1e46284_app1.pdf]

References
3. LaRoche KJ, Foster AM. “I kind of feel like sometimes I am shoving it under the carpet”: documenting women’s experiences with post-abortion support in Ontario. FACETS 2017 Sep 01;2(2):754-763 [FREE Full text] [doi: 10.1139/facets-2017-0059]


38. Sauro J. Measuring usability with the system usability scale (SUS). MeasuringU®. URL: https://measuringu.com/sus/ [accessed 2017-08-21]


Abbreviations

IUD: intrauterine device
mHealth: mobile health
SUS: system usability scale

©Roopan Gill, Gina Ogilvie, Wendy V Norman, Brian Fitzsimmons, Ciana Maher, Regina Renner. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.01.2024. 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.
Machine Learning for Early Prediction of Major Adverse Cardiovascular Events After First Percutaneous Coronary Intervention in Patients With Acute Myocardial Infarction: Retrospective Cohort Study

Pin Zhang¹,²*, MM; Lei Wu¹*, MM; Ting-Ting Zou¹, MM; ZiXuan Zou¹, MM; JiaXin Tu¹, MM; Ren Gong³*, PhD; Jie Kuang¹*, PhD

¹Jiangxi Provincial Key Laboratory of Preventive Medicine, School of Public Health, Nanchang University, Nanchang, China
²School of Public Health and Management, Nanchang Medical College, Nanchang, China
³Department of Cardiology, The Second Affiliated Hospital of Nanchang University, Nanchang, China
* these authors contributed equally

Corresponding Author:
Jie Kuang, PhD
Jiangxi Provincial Key Laboratory of Preventive Medicine
School of Public Health
Nanchang University
461 BaYi Road
Nanchang, 330006
China
Phone: 86 0791 8636228
Email: kuangjie@ncu.edu.cn

Abstract

Background: The incidence of major adverse cardiovascular events (MACEs) remains high in patients with acute myocardial infarction (AMI) who undergo percutaneous coronary intervention (PCI), and early prediction models to guide their clinical management are lacking.

Objective: This study aimed to develop machine learning–based early prediction models for MACEs in patients with newly diagnosed AMI who underwent PCI.

Methods: A total of 1531 patients with AMI who underwent PCI from January 2018 to December 2019 were enrolled in this consecutive cohort. The data comprised demographic characteristics, clinical investigations, laboratory tests, and disease-related events. Four machine learning models—artificial neural network (ANN), k-nearest neighbors, support vector machine, and random forest—were developed and compared with the logistic regression model. Our primary outcome was the model performance that predicted the MACEs, which was determined by accuracy, area under the receiver operating characteristic curve, and F1-score.

Results: In total, 1362 patients were successfully followed up. With a median follow-up of 25.9 months, the incidence of MACEs was 18.5% (252/1362). The area under the receiver operating characteristic curve of the ANN, random forest, k-nearest neighbors, support vector machine, and logistic regression models were 80.49%, 72.67%, 79.80%, 77.20%, and 71.77%, respectively. The top 5 predictors in the ANN model were left ventricular ejection fraction, the number of implanted stents, age, diabetes, and the number of vessels with coronary artery disease.

Conclusions: The ANN model showed good MACE prediction after PCI for patients with AMI. The use of machine learning–based prediction models may improve patient management and outcomes in clinical practice.

(JMIR Form Res 2024;8:e48487) doi:10.2196/48487

KEYWORDS
acute myocardial infarction; percutaneous coronary intervention; machine learning; early prediction; cardiovascular event
**Introduction**

Acute myocardial infarction (AMI) is a common clinical acute and severe disease with rapid onset, rapid progression, and high mortality [1-3]. In 2017, there were approximately 695,000 new cases of AMI in the United States, and it is estimated that 325,000 people will have recurrent events [4]. There are approximately 500,000 new cases of AMI in China every year, and 2.5 million patients have a history of myocardial infarction [5]. As technology has advanced, percutaneous coronary intervention (PCI) has become the primary approach for treating AMI. Although PCI can significantly reduce the fatality rate of AMI, the rate of major adverse cardiovascular events (MACEs) among patients after PCI is still very high, which seriously affects the clinical outcomes of patients [6-10]. A study by Copeland-Halperin et al [11] showed that the incidence of MACEs in patients with AMI one year after PCI was 17.8% [11].

Identifying patients with AMI undergoing PCI who are at high risk of MACEs may help clinical decision-making incorporate timely measures to improve clinical outcomes. Some studies, such as Global Registry of Acute Coronary Event [12], Thrombolysis in Myocardial Infarction Risk [13,14], and Acute Catheterization and Urgent Intervention Triage StrategY-PCI [15], as well as studies that generated the Mayo Clinic PCI Risk and the China Acute Myocardial Infarction scoring systems, have explored the risks after PCI [16]. Despite these advances, individualized prediction of MACEs remains challenging with low specificity and positive predictive accuracy, and most of the methods rely on traditional parameter models, such as logistic regression, to screen for variables and build a series of risk-scoring models.

In recent years, machine learning methods that rely on a strong self-learning capability, such as random forest (RF), k-nearest neighbors (KNN), support vector machine (SVM), and artificial neural network (ANN) have become increasingly prevalent in prognostic prediction [1,13,17,18]. By calling various functions, these models can extract and integrate information from all kinds of complex data to make better predictions. A study of a consecutive cohort of patients with hypertrophic cardiomyopathy (HCM) presented a machine learning–based model to identify individual patients with HCM at high risk of developing advanced heart failure symptoms. The results showed that the 5-year risk prediction of progressive heart failure in patients with HCM can be estimated [19].

We found that machine learning models, such as RF, ANN, SVM, and KNN, perform well in clinical prognosis prediction research. Thus, this study sought to develop a machine learning–based model, integrating clinical, anatomical, and laboratory features, to predict MACEs in patients who have recently been diagnosed with AMI after their first PCI and improve overall patient outcomes by implementing earlier management.

**Methods**

**Study Design, Setting and Participant Selection**

This retrospective cohort study was conducted at the Department of Cardiovascular Medicine, the Second Affiliated Hospital of Nanchang University (a teaching tertiary hospital), in Jiangxi Province, China. We collected electronic medical records of patients with AMI who underwent PCI for the first time from January 2018 to December 2019. These patients were followed up through December 2021.

The inclusion criteria of the participants were as follows:

- The patient was ≥18 years of age.
- This was the patient’s first clinically diagnosed AMI (clinical evidence of AMI as evident from the detection of a rise or fall of cardiac troponin values and at least one of the following symptoms of myocardial ischemia: symptoms of acute myocardial ischemia, new ischemic electrocardiogram (ECG) changes, and development of pathological Q waves.
- PCI was performed for the first time at this hospital.
- Among the left main artery, left circumflex branch, left anterior descending branch, and right coronary artery, at least one had stenosis ≥50%.
- Complete medical records and follow-up data were available.

The following exclusion criteria were applied:

- History of PCI and coronary artery bypass grafting treatment
- Complications from other heart diseases requiring surgical procedures, such as heart bypass
- Recent active bleeding
- An intracerebral mass or an aneurysm

We adopted the “Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research” to guide the reporting of our study [20].

**Data Collection, Definition of Outcomes, and Predictor Variables**

Data were collected from electronic health records, including demographic characteristics, clinical investigations, the first laboratory tests, and disease-related events. MACEs were defined as cardiomyopathies (excluding infectious, familial, alcohol, and drug-related cardiomyopathies), hypertensive heart disease, recurrent myocardial infarction, heart failure, sudden cardiac death, revascularization, malignant arrhythmia, and stent thrombosis [21]. Abnormal Q waves were identified by the clinician based on ECG results. Left ventricular ejection fraction (LVEF) was defined as normal (more than 50%), mildly abnormal (40% to 50%), moderately abnormal (30% to 40%), and severely abnormal (less than 30%) [22]. According to the number of diseased coronary vessels and implanted stents, they were classified as I, II, III, and IV.
**Ethics Approval**
This study was reviewed and approved by the Second Affiliated Hospital of Nanchang University Medical Ethics Committee (No. Review 2017 No. (098)).

**Data Preprocessing for Machine Learning Model Development**
All analyses were performed with R software (version 4.0.1; R Core Team). The patients were randomly assigned to training (n=953, 70%) and testing (n=409, 30%) data sets by calling the createDataPartition function using the random number method, and chi-square tests showed that there was no statistical difference between them ($\chi^2_{1}=2.169; P=0.14$). We developed machine learning models using the training data set. We analyzed the missing and out-of-range values with imputation methods. We used multiple imputation with chained equations to assign any missing predictor values [23]. The imputation processes were performed separately in the training and testing sets after the data were split. To improve the accuracy of the machine learning models and increase the speed of finding the optimal solution by gradient descent, we standardized and normalized all input variables before the model was built. To alleviate the problem of imbalanced classification samples, we adopted the random oversampling method. We used the ROSE package in R to generate new balanced training data. After random oversampling, the number of patients with MACE in the training data sets changed from 186 to 471.

**Predictor Selection for Model Development**
The model was built using demographic information (age and sex), personal comorbidities (diabetes and peripheral arterial disease), preoperative PCI (LVEF, the number of diseased vessels, and abnormal Q waves), serological examination (beta 2 microglobulin, B-type brain natriuretic peptide, glucose, serum creatinine clearance, and estimated glomerular filtration rate), and the characteristics of PCI (the number of implanted stents; n=65; Table S1 in Multimedia Appendix 1). A total of 12 variables with significant differences in the univariate analysis were included in the model development (Table 1).
Table 1. Baseline characteristics of the study patients (N=1362).

<table>
<thead>
<tr>
<th>Variables</th>
<th>MACE(^a) (n=252)</th>
<th>Non-MACE (n=1110)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;65</td>
<td>101 (40.08)</td>
<td>543 (48.92)</td>
<td>.04</td>
</tr>
<tr>
<td>65</td>
<td>94 (37.30)</td>
<td>332 (29.91)</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>57 (22.62)</td>
<td>235 (21.17)</td>
<td></td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Yes</td>
<td>75 (29.76)</td>
<td>261 (23.51)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>177 (70.24)</td>
<td>849 (76.49)</td>
<td></td>
</tr>
<tr>
<td>Vascular disease, n (%)</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Yes</td>
<td>111 (44.05)</td>
<td>569 (51.26)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>141 (55.95)</td>
<td>541 (48.74)</td>
<td></td>
</tr>
<tr>
<td>Abnormal Q wave, n (%)</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Yes</td>
<td>125 (49.60)</td>
<td>480 (43.24)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>127 (50.40)</td>
<td>630 (56.76)</td>
<td></td>
</tr>
<tr>
<td>LVEF(^b), n (%)</td>
<td></td>
<td></td>
<td>.005</td>
</tr>
<tr>
<td>&gt;50%</td>
<td>167 (66.27)</td>
<td>832 (74.95)</td>
<td></td>
</tr>
<tr>
<td>40%-50%</td>
<td>57 (22.62)</td>
<td>188 (16.94)</td>
<td></td>
</tr>
<tr>
<td>30%-40%</td>
<td>19 (7.54)</td>
<td>65 (5.86)</td>
<td></td>
</tr>
<tr>
<td>&lt;30%</td>
<td>9 (3.57)</td>
<td>25 (2.25)</td>
<td></td>
</tr>
<tr>
<td>Vessels with coronary artery disease, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>I</td>
<td>45 (17.86)</td>
<td>288 (25.95)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>75 (29.76)</td>
<td>370 (33.33)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>123 (48.81)</td>
<td>418 (37.66)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>9 (3.57)</td>
<td>34 (3.06)</td>
<td></td>
</tr>
<tr>
<td>Implanted stent number, n (%)</td>
<td></td>
<td></td>
<td>.004</td>
</tr>
<tr>
<td>No stent</td>
<td>10 (3.97)</td>
<td>40 (3.60)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>106 (42.06)</td>
<td>594 (53.51)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>84 (33.33)</td>
<td>301 (27.12)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>37 (14.68)</td>
<td>114 (10.27)</td>
<td></td>
</tr>
<tr>
<td>≥IV</td>
<td>15 (5.95)</td>
<td>61 (5.50)</td>
<td></td>
</tr>
<tr>
<td>Brain natriuretic peptide (pg/μL), mean (SD)</td>
<td>684.36 (997.90)</td>
<td>518.27 (773.65)</td>
<td>.01</td>
</tr>
<tr>
<td>Serum creatinine clearance (mL/min), mean (SD)</td>
<td>65.19 (30.18)</td>
<td>71.87 (44.35)</td>
<td>.02</td>
</tr>
<tr>
<td>EGFR(^c) (ml/min), mean (SD)</td>
<td>75.68 (28.92)</td>
<td>80.55 (31.82)</td>
<td>.03</td>
</tr>
<tr>
<td>Beta 2 microglobulin (mg/L), mean (SD)</td>
<td>3.23 (3.61)</td>
<td>2.72 (5.51)</td>
<td>.03</td>
</tr>
<tr>
<td>Glucose (mmol/L), mean (SD)</td>
<td>7.22 (3.32)</td>
<td>6.68 (3.00)</td>
<td>.02</td>
</tr>
</tbody>
</table>

\(^a\)MACE: major adverse cardiovascular events.
\(^b\)LVEF: left ventricular ejection fraction.
\(^c\)EGFR: estimated glomerular filtration rate.

**Model Testing and Performance Evaluation**

Based on a previous application of the model [24], the parameter range of the model was preset, and the GridSearchCV function was used to select the optimal parameters of each machine learning model.

To minimize potential overfitting in the above machine learning models, we called the trainControl function in the caret package of R language for 7-fold cross-validation during the development process. The model performance was assessed for accuracy, recall, precision, area under the receiver operating characteristic curve (AUC), and \(F_1\)-score in the testing data set. We identified
the important predictors through importance analysis of the variables. Logistic regression analysis was used to compare the absolute value of the coefficients of variables; RF was used to measure the importance of features by calculating information gain through entropy; and the ANN method was used to calculate the relative importance of variables based on the generalized weight method.

**Statistical Analysis**

The following R packages for machine learning approaches were used: caret, randomForest, and neuralnet. Baseline characteristics were compared with the Wilcoxon rank sum test for continuous variables and the chi-square test for categorical variables. We considered $P<.05$ (2-sided) to be statistically significant.

**Results**

A total of 1531 patients were screened; 140 patients who did not undergo PCI for the first time were excluded; 19 patients were lost to follow-up; and 1362 patients who were successfully followed up were included in this analysis (Figure 1). The mean follow-up time was 28.0 (SD 11.0) months (median 29.9 months). A total of 252 MACEs were observed, including 128 cases of recurrent myocardial ischemia and 117 cases of myocardial infarction and reinfarction. The positive rates of MACEs were 4.63%, 11.38%, 14.54%, and 18.50% at 30 days, 6 months, 1 year, and 3 years after PCI, respectively. MACEs occurred in 203 (18.7%) male patients and 49 (17.8%) female patients. As shown in Figure 2, the survival rate of the sample population decreased rapidly in the first 3 months after PCI, especially 30 days after PCI, and there was no difference in the log-rank test of the survival curve between male and female patients.

**Table 1** shows the baseline characteristics of the MACE group and the non-MACE group. Age, diabetes, peripheral and cerebrovascular history, LVEF, abnormal Q wave, the number of vessels with coronary artery disease, the number of implanted stents, brain natriuretic peptide, serum creatinine, estimated glomerular filtration rate, beta 2 microglobulin, and glucose were significantly different between the 2 groups ($P<.001$). The nonsignificant differences in variables between the 2 groups are shown in Table S1-S6 in Multimedia Appendix 1.

**Table 2** shows the performance of the 3 models with 7-fold cross-validation. ANN, KNN, SVM, RF, and logistic regression exhibited the best to worst performance in terms of their AUC, accuracy, recall, and $F_1$-score. However, KNN performed best in terms of precision. The average accuracy, recall, precision, AUC, and $F_1$-score of the ANN model were 80.52%, 81.33%, 69.94%, 83.68%, and 79.47%, respectively.

In the testing data set, the ANN model showed a higher AUC than RF and logistic regression. **Figure 3** shows that the AUCs of the ANN, RF, KNN, SVM, and logistic regression models were 0.805, 0.798, 0.772, 0.727, and 0.718, respectively; the average accuracy for the above 3 models was 0.821, 0.741, and 0.729, respectively, and the average $F_1$-scores were 0.804, 0.722, and 0.709, respectively.

The 10 most important predictors in the ANN model are shown in **Table 3**. These were LVEF (0.27), the number of implanted stents (0.14), age (0.13), diabetes (0.10), the number of vessels with coronary artery disease (0.09), vascular disease (0.08), brain natriuretic peptide (0.05), glucose (0.05), beta 2 microglobulin (0.04), and abnormal Q wave (0.02).

**Figure 1.** Flowchart for patient enrollment. AMI: acute myocardial infarction; MACE: major adverse cardiovascular event.

![Flowchart for patient enrollment](https://formative.jmir.org/2024/1/e48487)
Figure 2. Prognostic survival curve of patients with acute myocardial infarction undergoing percutaneous coronary intervention.

Table 2. Comparison of models for predicting major adverse cardiovascular events based on 7-fold cross-validation.

<table>
<thead>
<tr>
<th>Models</th>
<th>Accuracy, mean (SD)</th>
<th>Recall, mean (SD)</th>
<th>Precision, mean (SD)</th>
<th>AUC(^a), mean (SD)</th>
<th>(F_1)-score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression</td>
<td>72.37 (2.05)</td>
<td>67.33 (8.42)</td>
<td>59.62 (8.34)</td>
<td>73.52 (2.37)</td>
<td>71.11 (6.01)</td>
</tr>
<tr>
<td>K-nearest neighbors</td>
<td>81.44 (2.22)</td>
<td>80.23 (1.56)</td>
<td>70.22 (7.23)</td>
<td>81.87 (3.32)</td>
<td>77.95 (5.70)</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>74.91 (3.03)</td>
<td>80.03 (1.76)</td>
<td>65.94 (7.02)</td>
<td>78.68 (1.82)</td>
<td>76.41 (5.92)</td>
</tr>
<tr>
<td>Random forest</td>
<td>73.44 (1.58)</td>
<td>71.23 (1.56)</td>
<td>61.22 (7.23)</td>
<td>74.87 (2.12)</td>
<td>71.92 (6.30)</td>
</tr>
<tr>
<td>Artificial neural network</td>
<td>80.52 (1.13)</td>
<td>81.33 (0.56)</td>
<td>69.94 (7.02)</td>
<td>83.68 (1.82)</td>
<td>79.47 (4.57)</td>
</tr>
</tbody>
</table>

\(^a\)AUC: area under the receiver operating characteristic curve.

Figure 3. The area under the receiver operating characteristic (ROC) curve of artificial neural network (ANN), random forest (RF), k-nearest neighbors (KNN), support vector machine (SVM), and logistic regression models.
Table 3. Importance of each variable in the artificial neural network model.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left ventricular ejection fraction</td>
<td>0.27</td>
</tr>
<tr>
<td>The number of implanted stents</td>
<td>0.14</td>
</tr>
<tr>
<td>Age</td>
<td>0.13</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.10</td>
</tr>
<tr>
<td>The number of vessels with coronary artery disease</td>
<td>0.09</td>
</tr>
<tr>
<td>Vascular disease</td>
<td>0.08</td>
</tr>
<tr>
<td>Brain natriuretic peptide</td>
<td>0.05</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.05</td>
</tr>
<tr>
<td>Beta 2 microglobulin</td>
<td>0.04</td>
</tr>
<tr>
<td>Abnormal Q wave</td>
<td>0.02</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

In this study, we developed a machine learning–based model integrating clinical, anatomical, and laboratory test features to predict MACEs in patients with newly diagnosed AMI after their first PCI. The major findings suggest that the ANN model had higher predictive accuracy (accuracy of 87.99%, AUC of 0.81, and F1-score of 0.71), compared to RF, KNN, SVM, and logistic regression.

Among the patients with AMI in this study, the rates of MACEs at 30 days, 6 months, 1 year, and 3 years after PCI were 4.63%, 11.38%, 14.54%, and 18.50%, respectively, and the incidence of MACEs at 30 days after PCI was slightly less than the 5.5% reported in the Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction study (HORIZONS-AMI) [25]. The incidence of MACEs at half a year was higher than the 6.67% reported by Chow et al [26], consistent with the 2-year rate of MACEs reported by Sanmenxia City (18.06%). The survival condition of patients with AMI after PCI was slightly different from that in other studies. The participants in this study were all patients who were first diagnosed with AMI and underwent PCI for the first time, and their prognosis was better than that of patients with previous myocardial infarctions and multiple PCIs [27]. In addition, the progression of a patient’s disease is affected by not only individual differences but also access to medical resources and services. The HORIZONS-AMI trial was first reported in 2012. Although the treatment level in the HORIZONS-AMI trial was higher than that available in China at that time, with the development of China’s economy, the progress of science and technology, and the substantial improvement of medical care, the MACE rate obtained in our study was lower than that reported in the HORIZONS-AMI study.

One study found that machine learning demonstrated the highest performance for risk prediction in patients with extracardiac vascular disease for the prediction of both arrhythmogenic cardiomyopathy and MACEs [10]. McCord et al [28] proposed that machine learning can be used to assess AMI within 30 minutes and that the algorithm has high diagnostic and prognostic utility. In this study, 3 algorithms were used to predict MACE occurrence for patients with newly diagnosed AMI undergoing PCI treatment for the first time. The MACE prediction ability of the logistic regression model was lower than that of the ANN model and almost the same as that of the RF model. However, the positive predictive values of these 3 prediction models were not high. Kuang et al [29] also found that the ANN model had the best predictive value for the transition from mild cognitive impairment to Alzheimer disease with ideal stability [29]. The positive predictive values of the RF model and the logistic regression model were both approximately 50%, which means that their predictive ability for MACEs was poor. Their shortcomings may be associated with class imbalances [30], which can easily cause the predicted results to be biased toward a large number of classes (the positive type of fault can be placed into the negative class). ANNs, with their powerful self-adaptability, self-organization, fault tolerance, and “black box” operation of nonlinear mapping, are especially suitable for solving problems with complex internal mechanisms and have been widely used in various disciplines [31].

Our results indicated that the 3-year prognostic risk among patients with AMI undergoing their first PCI was mainly related to age, ECG characteristics, ventricular ejection ability, coronary artery lesions, stent implantation after PCI, and some serological variables. Yang et al [32] found that the risk ratio of hospital deaths after PCI was 3.723 (95% CI 2.86-4.84) for South Korean patients aged >65 years relative to those aged ≤65 years. A Korean multicenter AMI National Institutes of Health–registered project found that the MACE rate, 3 years after PCI, among patients with AMI with an LVEF <40% was 3.34 times that of the control group [33]. Fam et al [34] conducted a retrospective study on patients with clinical AMI in Asian multiethnic groups and found that the risk of MACEs among patients with diabetes, 2 years after PCI, was 1.84 times higher than that among patients without diabetes [34]. Diabetes is a chronic metabolic disease, and long-term diabetes is often accompanied by bleeding disorders, vascular endothelial dysfunction, small artery lesions, high blood sugar [35], hemostatic disorders [36], endothelial dysfunction, and a series of other changes [37]. These characteristics will accelerate the process of atherosclerotic disease deterioration. The number of coronary artery lesions
and the number of stents implanted in a patient are also positively correlated with the risk of postoperative MACEs to a certain extent. This may be because a higher number of vessels with coronary artery disease and the number of implanted stents tend to indicate a more serious condition, leading to a worse prognosis for the patients. Hongbo et al [38] found that the probability of a poor prognosis in patients with multiple coronary artery lesions was 20.0%, compared with 6.98% in patients with single coronary artery lesions [38].

The results of the machine learning model showed that predictors like LVEF, number of implanted stents, and age were more important to the model. LVEF is a common variable that reflects left ventricular function, and patients with a low LVEF have a significantly higher MACE rate [39]. An increase in age can lead to the aggravation of atherosclerosis [40]. The number of implanted stents may be related to the severity of the disease and the extent of the infarction [41]. This reminds us that we should pay special attention to the prognosis of patients with AMI who have a low LVEF value, older age, and more implanted stents in clinical practice.

Study Limitations
This study has some limitations. First, there may have been an issue of survival bias in the study, as patients with missing follow-up data were excluded. Second, the data have missing values. We have filled missing values with multiple imputation; however, imputation with these techniques could synthetically reduce the variance in these variables and may have affected the accuracy of the constructed model. Finally, although the models were internally validated with data from the same hospital, further work should include validation with external data from other hospitals or centers.

Conclusions
This study revealed that the ANN model showed good MACE prediction performance for patients with AMI after PCI, and it identified the most important predictors, which may aid in clinical decision-making and improve outcomes. This model needs to be externally validated in larger populations and multicenter settings.

Acknowledgments
The authors are grateful to the following individuals for contributing data: Li Li, Nan Wang, LiFang Deng, ShanLan Yang, and Bin Liu. This study was supported by the National Natural Science Foundation of China (grants 81960611, 82160645, 82360667, and 81260444), Natural Science Foundation of Jiangxi Province (grants 20192BAA208005 and 20212BAB206091), and the National Undergraduate Training Program for Innovation and Entrepreneurship (grant 202010403019).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Additional statistics.

References


Abbreviations

AMI: acute myocardial infarction
ANN: artificial neural network
AUC: area under the receiver operating characteristic curve
ECG: electrocardiogram
HCM: hypertrophic cardiomyopathy
HORIZONS-AMI: Harmonizing Outcomes with RevascularIZatiON and Stents in Acute Myocardial Infarction
KNN: k-nearest neighbors
LVEF: left ventricular ejection fraction
MACE: major adverse cardiovascular event
PCI: percutaneous coronary intervention
RF: random forest
SVM: support vector machine
Differences in Psychological Inflexibility Among Men With Erectile Dysfunction Younger and Older Than 40 Years: Web-Based Cross-Sectional Study

Junichi Saito1, PhD; Hiroaki Kumano2, MD, PhD; Mohammad Ghazizadeh3, MD, PhD; Chigusa Shimokawa3, MA; Hideki Tanemura3, MBA

1Comprehensive Research Organization, Waseda University, Saitama, Japan
2Faculty of Human Sciences, Waseda University, Saitama, Japan
3Logos Science Corp, Tokyo, Japan

Corresponding Author:
Junichi Saito, PhD
Comprehensive Research Organization
Waseda University
2-579-15, Mikajima, Tokorozawa
Saitama, 359-1192
Japan
Phone: 81 429498113
Email: tekuteke@aoni.waseda.jp

Abstract

Background: Psychological inflexibility is a core concept of acceptance and commitment therapy (ACT), which is a comprehensive, transdiagnostic interpretation of mental health symptoms. Erectile dysfunction (ED) is a condition that affects male sexual performance, involving the inability to achieve and maintain a penile erection sufficient for satisfactory sexual activity. Psychosocial factors primarily influence ED in men younger than 40 years, whereas biological factors are more likely to be the underlying cause in older men.

Objective: This web-based cross-sectional study examined differences in depression, anxiety, and psychological inflexibility among men with ED younger and older than 40 years in a Japanese population.

Methods: We used a web-based survey to gather data from various community samples. ED was assessed by the International Index of Erectile Function - 5 (IIEF-5) questionnaire, while depression, anxiety, and psychological inflexibility were evaluated by the Patient Health Questionnaire-9 (PHQ-9), General Anxiety Disorder-7 (GAD-7), Acceptance and Action Questionnaire-II (AAQ-II), Cognitive Fusion Questionnaire (CFQ), and Valuing Questionnaire–Obstacle Subscale (VQ-OB) questionnaires. The chi-square test estimated the scores of PHQ-9 and GAD-7 among men with ED, comparing those younger than 40 years and those older than 40 years. Additionally, a two-way ANOVA was conducted with ED severity and age group as independent variables, assessing psychological inflexibility.

Results: Valid responses from 643 individuals (mean age 36.19, SD 7.54 years) were obtained. Of these, 422 were younger than 40 years (mean age 31.76, SD 5.00 years), and 221 were older than 40 years (mean age 44.67, SD 2.88 years). There was a statistical difference in the prevalence of depression as judged by PHQ≥10 between men with ED younger and older than 40 years (P<.001). On the other hand, there was no difference in the prevalence of anxiety as judged by GAD≥10 (P=.12). The two-way ANOVA revealed that the interactions for CFQ (P=.04) and VQ-OB (P=.01) were significant. The simple main effect was that men with ED younger than 40 years had significantly higher CFQ (P=.01; d=0.62) and VQ-OB (P<.001; d=0.87) scores compared to those older than 40 years in moderate ED and severe ED. Additionally, it was found that men younger than 40 years with moderate to severe ED had significantly higher CFQ (P=.01; d=0.42) and VQ-OB (P=.02; d=0.38) scores compared to men younger than 40 years without ED. On the other hand, no interaction was found for AAQ-II (P=.16) scores.

Conclusions: To the best of our knowledge, this web-based cross-sectional study is the first to examine the relationship between psychological inflexibility and ED. We conclude that men with moderate and severe ED younger than 40 years have higher psychological inflexibility and might be eligible for ACT.

(JMIR Form Res 2024;8:e45998) doi:10.2196/45998

https://formative.jmir.org/2024/1/e45998

JMIR Form Res 2024 | vol. 8 | e45998 | p.80
(page number not for citation purposes)
KEYWORDS
erectile dysfunction; acceptance and commitment therapy; psychological inflexibility; depression; anxiety; men; cross-sectional study; psychological; utility; psychosocial; men; therapy; impotence; erection

Introduction
The efficacy of acceptance and commitment therapy (ACT) has been evaluated in numerous randomized controlled studies exploring various targeted conditions [1]. There is supporting evidence for ACT across various physical illnesses (eg, chronic pain [2], diabetes [3], epilepsy [4], cancer [5], and irritable bowel syndrome [6]). Many of these studies use a transdiagnostic method to analyze psychological issues within individual health conditions. Psychological inflexibility is a core concept of ACT, which is a comprehensive, transdiagnostic interpretation of mental health symptoms [7]. Psychological inflexibility highlights two interrelated processes: cognitive fusion and experiential avoidance. Cognitive fusion represents the phenomenon by which individuals are influenced by the literal meaning of their thoughts instead of viewing them as transient internal states [8]. Experiential avoidance represents an attempt or desire to suppress unwanted internal experiences, such as emotions, thoughts, memories, and bodily sensations [9]. These processes are obstacles to one’s valued living activities, decreasing well-being [10].

Erectile dysfunction (ED) is a condition that affects male sexual performance, involving the inability to achieve and maintain a penile erection sufficient for satisfactory sexual activity [11]. Several reviews and clinical guidelines are available for ED. However, many of these approaches to assessing and treating ED purely from a medical perspective seldom address the psychosocial components of ED [12]. Pharmacological treatment alone does not respond to all the concurrent factors of ED, including anxiety, loss of self-confidence, depressed mood, difficulties in a couple’s communication, relationship disputes, or a partner’s sexual problems [13]. Recent systematic reviews have shown that combining phosphodiesterase-5 inhibitors with psychological treatment exhibits significant potential for treating ED [14].

There is a widespread assumption that psychosocial factors primarily influence ED in men younger than 40 years, whereas biological factors are more likely to be the underlying cause of ED in older men. Moore et al [15] showed different symptom patterns among patients with ED according to age groups. They reported that younger men had comparatively more significant depressive symptoms, with lower relationship satisfaction, more negative reactions from partners, and lower job satisfaction. Given these findings, it is possible that people younger than 40 years are more psychologically inflexible than those older than 40 years and that ACT is more effective for them. However, studies on ACT for ED remain limited, with only a few identified. Therefore, this cross-sectional study used assessments to evaluate depression, anxiety, and psychological inflexibility in men younger and older than 40 years. However, it is well known that ED prevalence varies across geographical groups [16,17]; therefore, it is essential to research ED etiology according to different racial, cultural, religious, and socioeconomic backgrounds. There might be many potential patients with ED in Japan, so we conducted a web-based survey for this study.

Methods
Participants
To gather data from a wide range of community samples, we used a web-based survey, conducted with the assistance of a marketing research service provider (Rakuten Insight, Inc) in Japan. Based on the International Index of Erectile Function – 5 (IIEF-5) cutoff point [18], participants of all severities were recruited to include a certain percentage of patients of all ages. All enrolled participants followed the following criteria: (1) male; (2) aged 20 to 50 years; and (3) married or living with a fixed sexual partner for more than 6 months. The exclusion criteria were as follows: (1) sexual dysfunction caused by Peyronie disease or other organic lesions of the external genitalia; (2) prostate cancer, hypertensive disease, cardiac disease, cerebrovascular disease, chronic kidney disease, and diabetes; and (3) a history of sertraline or other medicines that may influence erection and psychological symptoms.

Participants were first instructed that this survey would be administered anonymously, and their responses were not compulsory. Then, those participants who agreed to participate in this research responded to the surveys. Participants were given points to exchange for items within the survey company’s system as a reward.

Ethical Considerations
This study was approved by the Waseda University Academic Research Ethical Review Committee (2019-363). The study protocol followed the guidelines for epidemiological studies in accordance with the Declaration of Helsinki.

Measurements
International Index of Erectile Function-5 (IIEF-5)
The Japanese version of IIEF-5 is a 5-item self-report questionnaire designed to measure erectile function [18]. Items are rated on a 5-point Likert-type scale, ranging from 1 to 5. The total score can range from 5 to 25, with high scores meaning high erectile function. Based on the original validation studies, the total score can then be interpreted as suggesting “no ED” (22-25), “mild ED” (17-21), “mild-to-moderate ED” (12-16), “moderate ED” (8-11), and “severe ED” (5-7).

Patient Health Questionnaire-9 (PHQ-9)
The Japanese version of the Patient Health Questionnaire-9 (PHQ-9) is a 9-item self-report questionnaire designed to measure depression [19]. Items are rated on a 4-point Likert-type scale, ranging from 0 to 3. The total score can range from 0 to 27, with high scores meaning high depression. Based on the original validation studies, the total score can then be interpreted as suggesting no depression (0-4), mild depression (5-9), moderate depression (10-14), moderately severe depression.
(15-19), or severe depression (20-27). A cutoff score of 10 is suggested as indicating a possible diagnosis of depressive disorder.

**Generalized Anxiety Disorder-7 (GAD-7)**

The Japanese version of the Generalized Anxiety Disorder-7 (GAD-7) questionnaire is a 7-item self-report questionnaire designed to measure generalized anxiety disorder [20]. Items are rated on a 4-point Likert-type scale, ranging from 0 to 3. The total score can range from 0 to 21, with high scores meaning high anxiety. Based on the original validation studies, the total score can then be interpreted as suggesting no anxiety (0-4), mild anxiety (5-9), moderate anxiety (10-14), or severe anxiety (14-21). A cutoff score of 10 is suggested as indicating a possible diagnosis of generalized anxiety disorder.

**Acceptance and Action Questionnaire-II (AAQ-II)**

The Japanese version of the Acceptance and Action Questionnaire-II (AAQ-II) is a 7-item self-report questionnaire designed to measure experiential avoidance [21]. Items are rated on a 7-point Likert-type scale, ranging from 1 to 7. The total score can range from 7 to 49, with high scores meaning high experiential avoidance.

**Cognitive Fusion Questionnaire (CFQ)**

The Japanese version of the Cognitive Fusion Questionnaire (CFQ) is a 7-item self-report questionnaire designed to measure cognitive fusion [22]. Items are rated on a 7-point Likert-type scale, ranging from 1 to 7. The total score can range from 7 to 49, with high scores meaning high cognitive fusion.

**Valuing Questionnaire–Obstacle Subscale (VQ-OB)**

The Japanese version of the Valuing Questionnaire–Obstacle Subscale (VQ-OB) is a 5-item self-report questionnaire designed to measure obstruction of valued living [23]. Items are rated on a 7-point Likert-type scale, ranging from 0 to 6. The total score can range from 0 to 30, with high scores meaning high obstruction of valued living.

**Statistical Analysis**

We used mean (SD) values to describe numerical data and counts and percentages to describe categorical data. The chi-square tests estimated categorical data, and numerical data were estimated by t tests. A two-way ANOVA test was used to assess the differences in men with ED aged younger and older than 40 years regarding psychological inflexibility and the interaction between them. Post hoc tests were conducted using the Holm method to control for type I errors. Cohen d index was calculated as effect sizes, serving as standardized indicators unaffected by sample sizes. All tests were 2-tailed, and a statistical difference was assumed when the P value was <.05. All statistical analyses were conducted through IBM SPSS Statistics (version 25.0; IBM Corp).

**Results**

We obtained valid responses from 643 individuals (mean age 36.19, SD 7.54 years). Of these, 422 were younger than 40 years (mean age 31.76, SD 5.00 years), and 221 were older than 40 years (mean age 44.67, SD 2.88 years). Table 1 shows the demographic characteristics of participants by age difference. No statistical difference was found in ED severity, phosphodiesterase-5 inhibitors use, and marriage status between men with ED younger and older than 40 years.

Table 1. Demographic characteristics of men with erectile dysfunction (ED) younger and older than 40 years.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>&lt;40 years of age (n=422)</th>
<th>≥40 years of age (n=221)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>31.76 (5.00)</td>
<td>44.67 (2.88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td>.44</td>
</tr>
<tr>
<td>Single</td>
<td>77 (18.25)</td>
<td>35 (15.84)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>345 (81.75)</td>
<td>186 (84.16)</td>
<td></td>
</tr>
<tr>
<td>Duration of marriage (years), mean (SD)</td>
<td>4.87 (3.90)</td>
<td>10.43 (7.53)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>IIEF-5 severity, n (%)</td>
<td></td>
<td></td>
<td>.84</td>
</tr>
<tr>
<td>No ED</td>
<td>84 (19.91)</td>
<td>49 (22.17)</td>
<td></td>
</tr>
<tr>
<td>Mild to mild-to-moderate ED</td>
<td>226 (53.55)</td>
<td>112 (50.68)</td>
<td></td>
</tr>
<tr>
<td>Moderate to severe ED</td>
<td>112 (26.54)</td>
<td>60 (27.15)</td>
<td></td>
</tr>
<tr>
<td>PDE-5 use, n (%)</td>
<td></td>
<td></td>
<td>.63</td>
</tr>
<tr>
<td>Not using</td>
<td>337 (79.86)</td>
<td>180 (81.45)</td>
<td></td>
</tr>
<tr>
<td>Using</td>
<td>85 (20.14)</td>
<td>41 (18.55)</td>
<td></td>
</tr>
</tbody>
</table>

aIIEF-5: International Index of Erectile Function - 5.
bPDE-5: phosphodiesterase-5 inhibitor.

The prevalence of depression as judged by PHQ≥10 among men younger than 40 years was 39.81% (168/422), and it was 24.89% (55/221) among those older than 40 years. There was a statistical difference in the prevalence of depression between the two groups (P<.001). In addition, the prevalence of anxiety, as judged by GAD≥10, was 27.25% (115/422) among men...
younger than 40 years, and it was 21.72% (48/221) among those older than 40 years. There was no difference in the prevalence of anxiety between men with ED in the two age groups ($P = .12$). Table 2 illustrates these results.

Table 2. Prevalence of depression and anxiety among men with erectile dysfunction younger and older than 40 years.

<table>
<thead>
<tr>
<th>Questionnaires and characteristics</th>
<th>&lt;40 years of age (n=422)</th>
<th>≥40 years of age (n=221)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHQ-9</strong>a</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No depression or mild depression</td>
<td>254 (60.19)</td>
<td>166 (75.11)</td>
<td></td>
</tr>
<tr>
<td>(PHQ-9&lt;10), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevalence of depression (PHQ-9≥10)</td>
<td>168 (39.81)</td>
<td>55 (24.89)</td>
<td></td>
</tr>
<tr>
<td><strong>GAD-7</strong>b</td>
<td></td>
<td></td>
<td>.13</td>
</tr>
<tr>
<td>No anxiety or mild anxiety (GAD-7&lt;10), n (%)</td>
<td>307 (72.75)</td>
<td>173 (78.52)</td>
<td></td>
</tr>
<tr>
<td>Prevalence of anxiety (GAD-7≥10), n (%)</td>
<td>115 (27.25)</td>
<td>48 (21.72)</td>
<td></td>
</tr>
</tbody>
</table>

$^a$PHQ-9: Patient Health Questionnaire-9.
$^b$GAD-7: Generalized Anxiety Disorder-7.

The two-way ANOVA was performed with ED severity and age (<40 or >40 years) as independent variables and the scores of AAQ, CFQ, and VQ-OB as dependent variables. The results showed no significant differences in AAQ-II ($P = .14$), CFQ ($P = .08$), and VQ-OB ($P = .30$) scores attributed to ED severity. Moreover, no difference in ED severity or psychological inflexibility depending on the duration of the marriage was found. On the other hand, there were significant differences in the scores of CFQ ($P = .04$) and VQ-OB ($P = .004$) attributed to age. As the interactions were significant for CFQ ($P = .04$) and VQ-OB ($P = .01$) scores, the simple main effect was examined. It was found that men with ED younger than 40 years had significantly higher CFQ ($P = .01$; $d = 0.62$) and VQ-OB ($P < .001$; $d = 0.87$) scores compared to those older than 40 years, in cases of moderate and severe ED. Additionally, it was found that men with moderate to severe ED younger than 40 years had significantly higher CFQ ($P = .01$; $d = 0.42$) and VQ-OB ($P = .02$; $d = 0.38$) scores compared to men with no ED younger than 40 years. These results are illustrated in Table 3 and Figures 1 and 2.

Table 3. Two-way ANOVA results of the influence of erectile dysfunction (ED) severity, age, and interaction on psychological inflexibility.

<table>
<thead>
<tr>
<th>Parameters and factors</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>$F$ test (df)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AAQ-II</strong>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED severity</td>
<td>281.60</td>
<td>140.80</td>
<td>1.91 (2,637)</td>
<td>.14</td>
</tr>
<tr>
<td>Age</td>
<td>204.74</td>
<td>204.74</td>
<td>2.78 (1,637)</td>
<td>.10</td>
</tr>
<tr>
<td>ED severity × age</td>
<td>267.27</td>
<td>133.64</td>
<td>1.81 (2,637)</td>
<td>.16</td>
</tr>
<tr>
<td><strong>CFQ</strong>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED severity</td>
<td>433.63</td>
<td>216.82</td>
<td>2.49 (2,637)</td>
<td>.08</td>
</tr>
<tr>
<td>Age</td>
<td>348.12</td>
<td>348.12</td>
<td>3.99 (1,637)</td>
<td>.04</td>
</tr>
<tr>
<td>ED severity × age</td>
<td>534.92</td>
<td>267.46</td>
<td>3.07 (2,637)</td>
<td>.04</td>
</tr>
<tr>
<td><strong>VQ-OB</strong>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ED severity</td>
<td>74.56</td>
<td>37.28</td>
<td>1.19 (2,637)</td>
<td>.30</td>
</tr>
<tr>
<td>Age</td>
<td>263.36</td>
<td>263.36</td>
<td>8.38 (1,637)</td>
<td>.004</td>
</tr>
<tr>
<td>ED severity × age</td>
<td>250.82</td>
<td>125.41</td>
<td>3.99 (2,637)</td>
<td>.01</td>
</tr>
</tbody>
</table>

$^a$AAQ-II: Acceptance and Action Questionnaire-II.
$^b$CFQ: Cognitive Fusion Questionnaire.
$^c$VQ-OB: Valuing Questionnaire–Obstacle Subscale.
Figure 1. Results of the two-way ANOVA and the simple main effect of the Cognitive Fusion Questionnaire (CFQ). ED: erectile dysfunction. \( P=.01 \) for "a," "b," and "c".

Figure 2. Results of the two-way ANOVA and the simple main effect of the Valuing Questionnaire–Obstacle Subscale (VQ-OB). ED: erectile dysfunction. \( P<.001 \) for "a" and \( P=.02 \) for "b".

Discussion

Principal Findings

This cross-sectional study evaluated depression, anxiety, and psychological inflexibility in men younger and older than 40 years with ED. There was no statistical difference in demographic characteristics between the two groups. The average age of the participants was 31.76 (SD 5.00) years in men younger than 40 years; the population was assumed to have mainly mild-to-moderate psychogenic ED. On the other hand, the average age of the participants was 44.67 (SD 2.88) years in men older than 40 years; the population was assumed to have mostly mild-to-moderate organic ED.

Depression was found in both groups. The results of our study were consistent with a previous study [15], which found that younger men had comparatively greater depressive symptoms. In contrast, the prevalence of anxiety was not different between the two age groups. One possible reason is that the anxiety in men with ED is not general anxiety but specific anxiety about sexual situations. Masters and Johnson [24] highlighted the
central role of sexual performance anxiety in couples presenting with sexual dysfunction [24]. In treating sexual dysfunctions, Kaplan [25] emphasizes the importance of addressing specific sources of sexual anxiety, such as fear of failure and not pleasing one’s partner [25]. Although a Japanese version does not exist now, it may be necessary to use a questionnaire like the Erectile Performance Anxiety Index [26].

Men with ED younger than 40 years had significantly higher CFQ and VQ-OB scores than those older than 40 years in cases of moderate and severe ED. Furthermore, men with moderate-to-severe ED younger than 40 years had significantly higher CFQ and VQ-OB scores compared to men without ED. These results partly support our hypothesis that men younger than 40 years are more psychologically inflexible than those older than 40 years. Cognitive fusion might be the critical component of ACT for ED. For example, the fusion with sexual performance anxiety, such as “I might fail again,” makes it impossible to pay attention to the sexual partner, which results in erectile failure. It is also consistent with Barlow’s theory [27]. Barlow [27] proposed a model for the interaction of anxiety and cognitive interference. This model examines how anxiety and cognitive interference interact, particularly in a sexual context, where a lack of control over one’s arousal diverts attention from erotic arousal to physical arousal and the negative consequences associated with failure to attain an erection.

On the other hand, there were no significant differences in the scores of the AAQ-II, which might be related to psychometric issues with AAQ-II. To date, the most used self-report measure of psychological inflexibility, especially experience avoidance, has been the AAQ-II. There was no significant difference in ED severity and psychological inflexibility depending on the duration of the marriage. However, various issues regarding the AAQ-II have emerged from the existing literature [28]. The authors found that the AAQ-II faced challenges in distinguishing distress (like negative affect and neuroticism) from experiential avoidance. For clinical application, researchers have expanded the range of measures for psychological inflexibility. They have developed specific versions of the AAQ-II tailored to different populations or disorders, with currently over 20 available versions (examples include those for the workplace, tinnitus, irritable bowel syndrome, exercise, and epilepsy). The disorder-specific AAQ-II variants indicate greater incremental validity in their targeted areas than the general AAQ-II [29]. Thus, developing a questionnaire on ED-related psychological inflexibility might be necessary.

There are some limitations to this study. First, this study used a cross-sectional approach, indicating merely “associations” rather than “causality” between psychological inflexibility and ED. Further controlled experimental and longitudinal studies are essential to delve deeper into the impact of psychological inflexibility on ED. Second, in this study, no responses were obtained from the partners of men with ED. Including the partners in the assessment and treatment of ED is recommended. It is desirable to obtain responses from partners in future studies. Finally, the specific racial or ethnic and socioeconomic profiles of the participants may restrict the broader applicability of the findings. The study was also conducted during the COVID-19 epidemic, which may have influenced the results.

Conclusions
To the best of our knowledge, this web-based cross-sectional study was the first to examine the relationship between psychological inflexibility and ED. We conclude that men with moderate and severe ED younger than 40 years have higher psychological inflexibility and might be eligible for the ACT. In addition, developing a Japanese version of the questionnaire is necessary to measure ED-related anxiety and psychological inflexibility.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to the Waseda University Academic Research Ethical Review Committee’s data-sharing policy but are available from the corresponding author upon reasonable request.

Conflicts of Interest
This research was funded by Logos Science Corp, Ltd, Tokyo, Japan. MG, CS, and HT are members of the Logos Science Corp, Ltd.

References

https://formative.jmir.org/2024/1/e45998 JMIR Form Res 2024 | vol. 8 | e45998 | p.85 (page number not for citation purposes)


Abbreviations

AAQ-II: Acceptance and Action Questionnaire - II
ACT: Acceptance and Commitment Therapy
CFQ: Cognitive Fusion Questionnaire
ED: Erectile Dysfunction
GAD-7: Generalized Anxiety Disorder - 7
IIEF-5: International Index of Erectile Function - 5
PHQ-9: Patient Health Questionnaire - 9
VQ-OB: Valuing Questionnaire–Obstacle Subscale
Determining Distinct Suicide Attempts From Recurrent Electronic Health Record Codes: Classification Study

Kate H Bentley¹²*, PhD; Emily M Madsen¹³*, BSc; Eugene Song¹³, BA; Yu Zhou¹³, MPP; Victor Castro⁴, MS; Hyunjoon Lee¹³, MSc; Younga H Lee²³, PhD; Jordan W Smoller¹²³, MD, SCD

¹Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States
²Department of Psychiatry, Harvard Medical School, Boston, MA, United States
³Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, United States
⁴Mass General Brigham Research Information Science and Computing, Somerville, MA, United States
*these authors contributed equally

Corresponding Author:
Kate H Bentley, PhD
Center for Precision Psychiatry
Department of Psychiatry
Massachusetts General Hospital
185 Cambridge Street
2nd Floor
Boston, MA, 02114
United States
Phone: 1 6177247741
Email: kbentley@mgh.harvard.edu

Abstract

Background: Prior suicide attempts are a relatively strong risk factor for future suicide attempts. There is growing interest in using longitudinal electronic health record (EHR) data to derive statistical risk prediction models for future suicide attempts and other suicidal behavior outcomes. However, model performance may be inflated by a largely unrecognized form of “data leakage” during model training: diagnostic codes for suicide attempt outcomes may refer to prior attempts that are also included in the model as predictors.

Objective: We aimed to develop an automated rule for determining when documented suicide attempt diagnostic codes identify distinct suicide attempt events.

Methods: From a large health care system’s EHR, we randomly sampled suicide attempt codes for 300 patients with at least one pair of suicide attempt codes documented at least one but no more than 90 days apart. Supervised chart reviewers assigned the clinical settings (ie, emergency department [ED] versus non-ED), methods of suicide attempt, and intercode interval (number of days). The probability (or positive predictive value) that the second suicide attempt code in a given pair of codes referred to a distinct suicide attempt event from its preceding code was calculated by clinical setting, method, and intercode interval.

Results: Of 1015 code pairs reviewed, 835 (82.3%) were nonindependent (ie, the 2 codes referred to the same suicide attempt event). When the second code in a pair was documented in a clinical setting other than the ED, it represented a distinct suicide attempt 3.3% of the time. The more time elapsed between codes, the more likely the second code in a pair referred to a distinct suicide attempt event from its preceding code. Code pairs in which the second suicide attempt code was assigned in an ED at least 5 days after its preceding suicide attempt code had a positive predictive value of 0.90.

Conclusions: EHR-based suicide risk prediction models that include International Classification of Diseases codes for prior suicide attempts as a predictor may be highly susceptible to bias due to data leakage in model training. We derived a simple rule to distinguish codes that reflect new, independent suicide attempts: suicide attempt codes documented in an ED setting at least 5 days after its preceding suicide attempt code had a positive predictive value of 0.90. This rule has the potential to minimize upward bias in model performance when prior suicide attempts are included as predictors in EHR-based suicide risk prediction models.
**Introduction**

Suicide is the tenth leading cause of death in the United States, with more than 48,000 suicide deaths annually [1]. Over the past 20 years, the suicide rate has increased by over 35% [2]. Most people who die by suicide have recently interacted with the health care system, with over half having a health care visit in the month prior to death [3,4]. Health care systems thus offer a key opportunity to identify people at high risk for suicide. Unfortunately, clinicians are poor at predicting who will make a suicide attempt [5] and traditionally studied risk factors perform no better than chance at predicting future suicidal behavior [6].

Recent work has focused on developing and validating machine learning models that use routinely collected electronic health record (EHR) data to predict future suicidal behavior [7]. Such models have demonstrated high levels of accuracy, exceeding that seen with clinician prediction and usual clinical risk factors [8-10]. EHR-based suicide risk prediction models, however, face one significant challenge that to date has not been adequately addressed. Suicide attempt is generally the outcome of interest in these models and is typically defined by International Classification of Diseases (ICD) diagnostic codes [11,12]. Within a given patient’s EHR, a suicide attempt code may be given multiple times across distinct health care encounters, often over very short periods of time (eg, days and weeks). Such “recurrent” codes may represent either distinct, new events (ie, multiple suicide attempts) or refer to the same event (ie, a single suicide attempt). The latter may occur when, for example, after making a suicide attempt, a patient has an emergency department (ED) visit followed by an inpatient hospitalization or outpatient follow-up encounters, with one (or multiple) suicide attempt codes assigned at each. In the absence of manual reviews of the narrative notes within patients’ EHRs, which cannot be performed at scale, it can be challenging to determine whether such recurrent suicide attempt codes, especially when documented over short time periods, refer to independent, distinct suicide attempts. Failure to make this important distinction can result in a form of “data leakage” in which the outcome to be predicted is included among features used for the prediction. This can result in substantial inflation of model performance [13].

To address this issue, some researchers have restricted model development to predict only the first occurrence of a suicide attempt code in a patient’s EHR [14-16]. This approach has a major limitation, however, in that a past suicide attempt is among the strongest known predictors of future suicidal behavior [17]. Thus, models that predict only the first documented suicide attempt ignore the subset of patients who may be at highest risk and thus of greatest clinical concern: those with a prior suicide attempt. Another approach is to include any previous suicide attempt codes as predictors of a subsequent suicide attempt code [18-21] thus including potential “repeat attempters” in these models. This approach, however, poses a significant risk of artificially inflating model performance if subsequent codes do not in fact refer to new suicide attempts. In other words, if a suicide attempt code instance used as an outcome actually indexes an attempt that was included a predictor, model performance will be inflated.

To minimize the risk of data leakage while retaining the option of including prior attempts as predictors, we aimed to develop an automated rule for determining whether recurrent suicide attempt codes in the EHR refer to distinct events. Such a rule might be based on relevant variables including clinical setting (eg, a suicide attempt code documented in the ED may be more likely to refer to a new suicide attempt event than one given in a non-ED setting), method (eg, suicide attempt codes that specify different methods may be more likely to refer to distinct events than codes specifying the same method), and time (eg, the more time elapsed between 2 suicide attempt codes, the less likely it may be that the codes refer to the same event). Here, we conducted a comprehensive manual EHR chart review to derive an automated rule that could identify criteria for selecting distinct suicide attempts with high confidence.

**Methods**

**Data Source**

The data source for this study was the Mass General Brigham (MGB) Research Patient Data Registry [22]. This registry covers 6.7 million patients treated in MGB-affiliated hospitals including the Massachusetts General Hospital and Brigham and Women’s Hospital in Boston.

**Ethics Approval**

This research was approved by the MGB institutional review board, which granted a waiver of informed consent (protocol #2018P0001508).

**Case Definition and Inclusion Criteria**

Details of the development of our EHR-based case definitions for suicide attempt in the MGB health care system are reported elsewhere [14,15]. In brief, we first identified candidate ICD, Ninth Revision (ICD-9) and ICD, Tenth Revision (ICD-10) codes that are likely to capture suicide attempts. Next, expert clinicians conducted manual chart reviews of 670 patients (over 3000 narrative notes) to determine a final set of codes that capture suicide attempts with a positive predictive value (PPV) of >0.70: for ICD-9, E95*, 965*, 967*, 969*, and 881*, and for ICD-10, X71*-X83*, T14.91*, T36*-T50* where the sixth character is 2 (except for T36.9, T37.9, T41.4, T42.7, T43.9, T45.9, T47.9, and T49.9, where the fifth is 2) and T51*-T65* where the sixth character is 2 (except for T36.9, T37.9, T41.4, T42.7, T43.9, T45.9, T47.9, and T49.9, where the fifth is 2) and T51*-T65* where the sixth character is 2 (except for T36.9, T37.9, T41.4, T42.7, T43.9, T45.9, T47.9, and T49.9, where the fifth is 2) and T51*-T65* where the sixth character is 2 (except for T36.9, T37.9, T41.4, T42.7, T43.9, T45.9, T47.9, and T49.9, where the fifth is 2) and T51*-T65* where the sixth character is 2 (except for T36.9, T37.9, T41.4, T42.7, T43.9, T45.9, T47.9, and T49.9, where the fifth is 2) and T51*-T65* where the sixth character is 2 (except for T36.9, T37.9, T41.4, T42.7, T43.9, T45.9, T47.9, and T49.9, where the fifth is 2).
T61.9, T62.9, T63.9, T64.0, T64.8, and T65.9, where the fifth is 2).

For this study, we randomly selected a sample of 300 patients with 2 suicide attempt codes documented at least one but no more than 90 days apart (the “narrow sample”). This interval was chosen to capture codes that were given within a narrow time frame and thus potentially enriched for being “leaked” codes. In a sensitivity analysis, we randomly selected a second, smaller sample of 100 patients with 2 suicide attempt codes documented at least 1 day apart but with no other restrictions on intercode interval (the “broad sample”). A total of 31 patients appeared in both narrow and broad samples. Patients for whom we were unable to confidently locate the narrative notes corresponding to documented suicide attempt codes (eg, no narrative notes available within 30 days of the suicide attempt code date, narrative notes recorded on paper and never migrated to the EHR) were excluded after the sampling process.

Procedure
Under the supervision of JWS (a senior clinician with expertise in the treatment of suicidal behavior), 2 study team members (EMM and ES) manually reviewed the EHR clinical encounter data (including narrative notes) relevant to each pair of suicide attempt codes (“code pair”) per sampled patient (1015 in the narrow sample and 300 in the broad sample; 1253 unique codes across the 2 samples). Each code pair comprised a given suicide attempt code and the immediately (temporally) preceding code in a patient’s EHR. All applicable code pairs per patient were examined (including other code pairs with >90-day intervals for patients in the narrow sample). Chart reviewers assigned the following variables to each code pair: (1) whether the code pair referred to 2 distinct suicide attempts (dichotomous variable indicating distinct or not distinct suicide attempts), (2) clinical setting in which each code in the pair was documented (dichotomous variable indicating ED or non-ED [eg, outpatient and inpatient setting]), (3) suicide attempt method of each code in the pair (categorical variable with 6 categories derived from previous literature: poisoning, cutting or piercing, hanging or strangulation or suffocation, jumping, firearm, and other [which included codes with no specified method]), and (4) time elapsed (in days) between codes in each pair [23]. When there were multiple encounters with suicide attempt codes on the same day, these variables were assigned to codes at the day level; see Table S1 in Multimedia Appendix 1 for an example of sampled codes (and the variables assigned to each code and code pair) for a deidentified patient.

Data Analysis
We defined PPV as the probability that the second code in a pair of codes identified a new suicide attempt independent of the first code in the pair. To mimic the approach that would likely be taken in building predictive models, each code pair was treated independently (ie, we did not account for the nested nature of code pairs within patients). First, for the narrow sample, we calculated in Excel [Microsoft] PPVs and 95% CIs by clinical setting, suicide attempt method, and intercode interval, respectively. For clinical setting, we calculated the PPVs for 4 possible code pair types: (1) both codes documented in the ED (ED/ED), (2) first code ED and second code non-ED (ED/non-ED), (3) first code non-ED and second code ED (non-ED/ED), and (4) neither code ED (non-ED/non-ED). For suicide attempt method, we calculated the PPVs of 2 possible code pair types: (1) same suicide attempt method for codes in a pair and (2) different suicide attempt methods for codes in a pair. For intercode interval, we first calculated PPVs for all 7-day intervals from 1 to 91 days, followed by collapsing across intervals from 92 days on. We then calculated the PPVs for time intervals within each of the 6 (4 clinical settings and 2 suicide attempt methods) code pair types. To derive our proposed rule, we set our benchmark PPV to 0.90. For each of the 6 code pair types, we determined the minimum time elapsed between codes (ie, interval floor) at which the PPV was at least 0.90. For a sensitivity analysis, we computed the same series of PPVs for the broad sample.

Results

Descriptive Statistics
The mean number of suicide attempt codes per patient in the narrow sample was 3.38 (SD 4.62; range 1–47). A total of 225 (75%) patients had <4 codes and 281 (93.7%) had <10 codes. A total of 210 (20.7%) code pairs had a second code reflecting a subsequent encounter for a condition for which the patient had received active treatment (indicated by a seventh “D” character).

Regarding how often the codes in a pair referred to distinct suicide attempts, of the 300 patients in the narrow sample, only 81 (27%) had more than one confirmed (by manual chart review) suicide attempt captured by the reviewed code pairs. Of the 1015 code pairs, only 180 (17.7%) referred to 2 distinct suicide attempt events. Table S1 in Multimedia Appendix 1 presents an example of sampled codes (and the variables assigned to each code and code pair) for a deidentified patient.

For clinical setting, the most common code pair types were non-ED/non-ED (n=542, 53.4%) followed by ED/ED (n=274, 27%). Regarding the 749 total non-ED codes, the most commonly represented clinical setting was inpatient (n=411, 54.9% of all non-ED codes), followed by other or unclear setting (n=149, 19.9% codes), intensive or critical care units (n=134, 17.9% codes), and outpatient (n=55, 7.3% codes). For suicide attempt method, the majority of code pairs (n=766, 75.5%) comprised 2 codes that referred to the same method. The median interval between codes in each code pair, across all codes, was 1 day. Among code pairs that referred to distinct suicide attempt events, the median interval was 35 days.

PPVs

Clinical Setting
Non-ED/ED code pairs (23 total code pairs) had the highest PPVs (0.96, 95% CI 0.87-1.04) for distinct suicide attempt events (Table 1). ED/ED pairs (274 total code pairs) had the second-highest PPVs (0.49, 95% CI 0.43-0.55). When the second code in a pair was assigned in a non-ED setting, PPVs were low (below 0.10).

In a sensitivity analysis, we excluded codes or encounters documented in inpatient settings with a prior code on the previous day from an inpatient or critical or intensive care
setting. For example, if a patient was given suicide attempt codes on three consecutive days in an inpatient setting, we only used the day 1 code. This resulted in 792 (versus 1015) analyzed code pairs. The results were overall very similar to when we did not exclude contiguous inpatient codes (Multimedia Appendix 2).

Table 1. Code pairs in the narrow sample defined by the clinical setting (ED\(^a\) or non-ED) of the first and second codes in each pair.

<table>
<thead>
<tr>
<th>First code clinical setting</th>
<th>Second code clinical setting</th>
<th>Code pairs (percentage of all code pairs), n (%)</th>
<th>Interval between codes (days), median (Q1(^b), Q3(^c))</th>
<th>Interval between codes (days), mean (SD)</th>
<th>Code pairs referring to distinct attempts, n</th>
<th>PPV(^d) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-ED</td>
<td>Non-ED</td>
<td>542 (53.4)</td>
<td>1 (1, 1)</td>
<td>3.58 (8.64)</td>
<td>14</td>
<td>0.03 (0.01-0.04)</td>
</tr>
<tr>
<td>ED</td>
<td>Non-ED</td>
<td>176 (17.3)</td>
<td>1 (1, 3)</td>
<td>6.47 (34.23)</td>
<td>10</td>
<td>0.06 (0.02-0.09)</td>
</tr>
<tr>
<td>Non-ED</td>
<td>ED</td>
<td>23 (2.3)</td>
<td>52 (13, 125)</td>
<td>154.09 (286.63)</td>
<td>22</td>
<td>0.96 (0.87-1.04)</td>
</tr>
<tr>
<td>ED</td>
<td>ED</td>
<td>274 (27)</td>
<td>5 (1, 36)</td>
<td>53.79 (211.77)</td>
<td>134</td>
<td>0.49 (0.43-0.55)</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>1015</td>
<td>1</td>
<td>21.05 (122.43)</td>
<td>180</td>
<td>0.18 (0.15-0.20)</td>
</tr>
</tbody>
</table>

\(^a\)ED: emergency department.  
\(^b\)Q1: first quartile. 
\(^c\)Q3: third quartile.  
\(^d\)PPV: positive predictive value.

**Suicide Attempt Method**

For suicide attempt method (same versus different method for 2 codes in a code pair), the PPVs were below 0.25 (Table 2). Table S6 in Multimedia Appendix 3 shows PPVs for each combination of the 6 aforementioned specific method categories derived from previous literature. All PPVs for strata containing more than 1 code pair were at or below 0.50.

Table 2. Code pairs defined by whether the first and second codes referred to the same or a different suicide attempt method.

<table>
<thead>
<tr>
<th>First and second code</th>
<th>Code pairs (percentage of all code pairs), n (%)</th>
<th>Code pairs referring to distinct attempts, n</th>
<th>PPV(^a) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same method</td>
<td>766 (75.5)</td>
<td>128</td>
<td>0.17 (0.14-0.19)</td>
</tr>
<tr>
<td>Different method</td>
<td>249 (24.5)</td>
<td>52</td>
<td>0.21 (0.16-0.26)</td>
</tr>
<tr>
<td>Overall</td>
<td>1015</td>
<td>180</td>
<td>0.18 (0.15-0.20)</td>
</tr>
</tbody>
</table>

\(^a\)PPV: positive predictive value.

**Intercode Interval**

Table 3 presents PPVs for code pairs broken down by 7-day (week-long) intervals; the majority (n=797, 78.5%) of code pairs had an intercode interval of 7 days or less. The more days elapsed between 2 codes, the larger the PPV (and, fewer code pairs per strata). Table S7 in Multimedia Appendix 4 presents PPVs for code pairs broken down by interval and clinical setting (non-ED/non-ED, ED/non-ED, non-ED/ED, ED/ED), and Table S8 in Multimedia Appendix 5 presents PPVs for code pairs broken down by interval and suicide attempt method (same versus different). In another sensitivity analysis, given that ICD-9 is no longer used, we also computed all PPVs reported in Tables 1-3 when excluding code pairs with at least one ICD-9 coded event. The same pattern of findings held, with 95% CIs for all PPVs overlapping with those in Tables 1-3.
Table 3. Code pairs defined by intercode interval.

<table>
<thead>
<tr>
<th>Intercode interval</th>
<th>Code pairs (percentage of all code pairs), n (%)</th>
<th>Code pairs referring to distinct attempts, n</th>
<th>PPV(^a) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-7 days</td>
<td>797 (78.5)</td>
<td>31</td>
<td>0.04 (0.03-0.05)</td>
</tr>
<tr>
<td>8-14 days</td>
<td>48 (4.7)</td>
<td>19</td>
<td>0.40 (0.26-0.53)</td>
</tr>
<tr>
<td>15-21 days</td>
<td>31 (3)</td>
<td>17</td>
<td>0.55 (0.37-0.72)</td>
</tr>
<tr>
<td>22-28 days</td>
<td>20 (2)</td>
<td>14</td>
<td>0.70 (0.50-0.90)</td>
</tr>
<tr>
<td>29-35 days</td>
<td>17 (1.7)</td>
<td>10</td>
<td>0.59 (0.35-0.82)</td>
</tr>
<tr>
<td>36-42 days</td>
<td>17 (1.7)</td>
<td>15</td>
<td>0.88 (0.73-1.04)</td>
</tr>
<tr>
<td>43-49 days</td>
<td>10 (1)</td>
<td>8</td>
<td>0.80 (0.55-1.05)</td>
</tr>
<tr>
<td>50-56 days</td>
<td>18 (1.8)</td>
<td>14</td>
<td>0.78 (0.59-0.97)</td>
</tr>
<tr>
<td>57-63 days</td>
<td>5 (4.9)</td>
<td>3</td>
<td>0.60 (0.17-1.03)</td>
</tr>
<tr>
<td>64-70 days</td>
<td>7 (0.7)</td>
<td>6</td>
<td>0.86 (0.60-1.12)</td>
</tr>
<tr>
<td>71-77 days</td>
<td>2 (0.2)</td>
<td>2</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>78-84 days</td>
<td>9 (0.9)</td>
<td>9</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>85-91 days</td>
<td>6 (0.6)</td>
<td>6</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>92+ days</td>
<td>28 (2.7)</td>
<td>26</td>
<td>0.93 (0.83-1.02)</td>
</tr>
<tr>
<td>Overall</td>
<td>1015</td>
<td>180</td>
<td>0.18 (0.15-0.20)</td>
</tr>
</tbody>
</table>

\(^a\)PPV: positive predictive value.

As shown in Figure 1, across all code pairs, pairs with an interval of at least 53 days had a PPV of 0.90 (range 0.88-0.93). The interval floors meeting our benchmark PPV (at least 0.90) within each of the 6 code pair types are also labeled in Figure 1 (clinical setting) and Figure 2 (suicide attempt method). For non-ED/ED code pairs (23 code pairs), an interval floor of 1 day had a PPV of 0.96. When both codes were assigned in the ED (271 code pairs), PPV reached 0.90 when the intercode interval was at least 5 days. When the second code in a pair was documented in an ED (regardless of the setting in which the first code was documented), PPV was 0.91 when the intercode interval was 5 days (the PPV was 0.89 for 4 days). Thus, whenever the second code in a pair was documented in an ED at least 5 days after the previous code, the probability that the second code referred to an independent suicide attempt was at least 90%.

Figure 1. PPVs for interval floors by code pair types defined by clinical setting. The labeled data points indicate the interval floor at which the PPV was at least 0.90 (or the maximum PPV). Gray lines reflect PPVs for interval floors across all code pair types. Red lines refer to code pairs documented in ED (first code) and ED (second code) settings (ED/ED). Blue lines are non-ED/ED code pairs; purple lines ED/non-ED; and green lines non-ED/non-ED. ED: emergency department; PPV: positive predictive value.
Figure 2. PPVs for interval floors by code pair types defined by suicide attempt method with intervals greater than or equal to the plotted interval floor values. The labeled data points indicate the interval floor at which the PPV was at least 0.90 (or the maximum PPV). Gray lines reflect PPVs for interval floors across all code pair types. Orange lines refer to code pairs in which the 2 codes refer to the same suicide attempt method. Blue lines refer to code pairs in codes in the pairs referring to different suicide attempt methods. PPV: positive predictive value.

Sensitivity Analysis: Broad Sample

Results from the same series of analyses in the broad sample are presented in Multimedia Appendix 6. Of the 100 patients, 45 (45%) had more than 1 confirmed suicide attempt. Of the 300 code pairs, 86 (28.7%) referred to 2 distinct suicide attempts. The median interval between codes in each pair was also 1 day. Among code pairs that referred to distinct suicide attempts, the median interval was 133 days. Overall, we found a similar pattern of PPVs (in almost all cases overlapping 95% CIs) to those from the narrow sample. Across all code pairs in the broad sample, those with an interval of at least 37 days had a PPV of 0.90 (range 0.87-0.93). When both codes were given in the ED (86 pairs), PPV reached 0.90 when the interval was at least 2 days.

Discussion

Primary Findings

Machine learning suicide risk prediction models that leverage routinely collected EHR data can outperform clinician assessment [8] and have the potential to improve how patients at risk for suicide are identified and treated. These models are typically trained using ICD codes to label suicide attempts. An under-appreciated challenge when building these models, however, is that ICD codes indexing a single suicide attempt are often used repeatedly across multiple encounters. This could create a substantial problem for models that incorporate prior suicide attempts, an established risk factor, in predicting subsequent attempts or suicidal behavior.

Some investigators side-step this issue by restricting model predictions to only the first occurrence of a suicide attempt code. This approach, however, limits the utility of prediction models by ignoring prior attempts, the best-known risk factor for suicidal behavior, and limiting their application to a subset of those at risk; prior studies indicate that nearly one-quarter of those who engage in deliberate self-harm have recurrent episodes within 3 years [24]. Here we aimed to develop a portable, automated rule for determining when recurrent suicide attempt codes refer to distinct suicide attempt events in a patient’s history. Based on chart review of clinical encounters corresponding to 1015 unique ICD code pairs, we found that, for patients with more than 1 documented suicide attempt code, repeat codes most often (>80% of the time) reflected nonindependent events, underscoring the high frequency of “leaked” suicide attempt codes. When collapsing across all clinical settings, repeat codes needed to be documented at least 53 days after the preceding code in order to refer (with probability >90%) to a new, distinct suicide attempt. However, repeat codes documented in an ED at least 5 days after the preceding suicide attempt code were likely (probability >90%) to refer to a new, distinct suicide attempt.

The most informative variables for determining whether recurrent suicide attempt codes referred to distinct suicide attempts were the clinical setting in which the codes were documented and the time elapsed between codes. First, regarding clinical setting, when a suicide attempt code was documented in an ED after the preceding code, it referred to a new suicide attempt more than half the time. Suicide attempt codes documented in non-ED settings, accounting for most of the second codes among all code pairs, however, were highly unlikely to refer to a new suicide attempt (probability <5%). This may be due to the fact that the vast majority (nearly three-quarters) of non-ED codes occurred in inpatient or intensive or critical care units, where patients may be treated over the course of several days or longer, potentially accumulating multiple suicide attempt codes that all refer to the same index event that may have prompted inpatient or intensive treatment. This pattern of findings, for one, highlights the considerable risk of treating all recurrent suicide attempt codes (especially those from non-ED settings) as distinct events, and the potential importance of using a simple rule, such as that proposed here, to identify probable distinct suicide attempt events.
Along these lines, the more time elapsed between 2 suicide attempt codes, the more likely it was the codes referred to distinct events. Combining these 2 variables—clinical setting and time elapsed—provided a simple rule for determining whether recurrent suicide attempt codes refer to distinct events with at least 90% probability. Although the accuracy of our proposed rule (at least 5 days elapsed between a code given in the ED and the preceding code) may differ in other health care systems, we recommend that others consider taking into account these 2 variables when incorporating recurrent suicide attempt codes in EHR-based suicide risk prediction models.

Perhaps surprisingly, whether the coded suicide attempt method for 2 codes in a pair was the same or different did not provide value in identifying distinct suicide attempt events. However, in the relatively small proportion of code pairs (24.5%) that referred to different methods, the most common “profile” was 1 code with a specific method (eg, poisoning and cutting or piercing) and the other code with method categorized as “other” (not a different specific method); notably, the “other” category included codes lacking any specified method. Thus, the fact that method did not help identify distinct events may largely reflect inconsistencies in how or whether the suicide attempt method is coded by providers. In contrast, neither of the other 2 variables examined (clinical setting nor intercode interval) should be impacted by irregular coding practices, and thus may also be more scalable and reliable for other health care systems planning to use this or a similar rule.

Our derived rule (at least 5 days elapsed between a code from the ED and the preceding code) may have more impact on certain suicide-related prediction tasks than others. For example, it may be especially relevant when estimating patients’ risk of repeat suicidal behavior, for example after an ED visit for suicidal behavior, which could influence clinical decision-making at the point of care (eg, about discharge home or to outpatient care versus hospitalization). This rule may have less impact for other related prediction tasks, such as estimating patients’ risk of suicidal behavior after nonsuicide-related outpatient visits or broader population-based prediction efforts [25]. These results may also be less relevant for models that solely predict fatal self-harm or suicide deaths [26,27]. Future work should systematically evaluate the performance and clinical utility of models that do and do not incorporate the proposed rule for incorporating recurrent suicide attempt codes across a range of prediction goals and clinical contexts.

Our results must be considered in the context of a few key limitations. First, some of the sampled patients may have presented to hospitals outside of the MGB system for suicide attempts. In these cases, the corresponding diagnostic codes and contextual information were either unavailable or only sporadically recorded in narrative notes at subsequent clinical encounters within MGB. We also excluded sampled patients for whom chart reviewers could not confidently match data pulled from the MGB Research Patient Data Registry to the narrative notes.

**Conclusions**

This analysis indicates that EHR-based suicide attempt prediction models that include ICD codes for prior attempts as a predictor may be highly susceptible to bias due to data leakage in model training. Our proposed rule for circumventing this issue should minimize this bias and its inflationary effect on model performance metrics. The key variables included in our rule (clinical setting and time elapsed between codes) are widely available in health system data warehouses and should be easily integrated into EHR-based models. It is also possible that the approach taken in this study may be relevant for developing and refining machine learning models aimed to predict other episodic events of interest that can be repeatedly documented in the health record, such as unintentional overdose, domestic abuse, or episodes of violence. If effectively implemented into existing and future suicide risk prediction models, this rule could increase the robustness and validity of machine-learning based approaches to identifying the individuals at highest risk for suicide, and ultimately advance suicide prevention efforts in health care contexts on a large scale.

**Acknowledgments**

The authors thank the Enterprise Research Infrastructure and Services at MGB for their in-depth support and for the provision of the research patient data registry and the ERISOne Linux cluster. This work was supported by grants (NIMH R01 MH117599; JWS) and (K23MH120436; KHB) from the National Institute of Mental Health, and a gift from the Tommy Fuss Fund (JWS). JWS is a member of the Leon Levy Foundation Neuroscience Advisory Board, the Scientific Advisory Board of Sensorium Therapeutics (with equity), and has received an honorarium for an internal seminar Tempus Labs. He is principal investigator of the research patient data registry and the ERISOne Linux cluster. This work was supported by grants (NIMH R01 MH117599; K23MH120436; KHB) from the National Institute of Mental Health, and a gift from the Tommy Fuss Fund (JWS). The authors thank the Enterprise Research Infrastructure and Services at MGB for their in-depth support and for the provision of the research patient data registry and the ERISOne Linux cluster. This work was supported by grants (NIMH R01 MH117599; JWS) and (K23MH120436; KHB) from the National Institute of Mental Health, and a gift from the Tommy Fuss Fund (JWS). JWS is a member of the Leon Levy Foundation Neuroscience Advisory Board, the Scientific Advisory Board of Sensorium Therapeutics (with equity), and has received an honorarium for an internal seminar Tempus Labs. He is principal investigator of the research patient data registry and the ERISOne Linux cluster. This work was supported by grants (NIMH R01 MH117599; K23MH120436; KHB) from the National Institute of Mental Health, and a gift from the Tommy Fuss Fund (JWS).

**Data Availability**

The data used in this study cannot be made publicly available due to restrictions relating to the use of EHR data.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Sampled codes or code pairs and designations per manual chart review for an example (deidentified) patient.
Multimedia Appendix 2
Sensitivity analysis (excluding contiguous codes from inpatient settings).

Multimedia Appendix 3
Code pairs in the narrow sample defined by specific category of suicide attempt method (poisoning, cutting or piercing, hanging or strangulation or suffocation, jumping, firearm, or other) of the first and second code in each code pair.

Multimedia Appendix 4
Code pairs in the narrow sample defined by both the clinical setting (ED or non-ED) of and the interval (in days) between the first and second codes in each pair.

Multimedia Appendix 5
Code pairs in the narrow sample defined by both suicide attempt method (same or different) and the interval (in days) between first and second codes in each pair.

Multimedia Appendix 6
Sensitivity analysis (results for broad sample).

References


Abbreviations
EN: emergency department  
EHR: electronic health record  
ICD: International Classification of Diseases  
MGB: Mass General Brigham  
PPV: positive predictive value